The Hotel Test of multitasking: Psychometric properties and norms

Andriana Milia-Argeiti (B.A. Hons, MSc)

Submitted in partial fulfilment of the requirements for the degree of

Doctorate in Clinical Psychology

Institute of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

May 2022

Table of contents

List of tables

List of figures

Acknowledgments

First, I would like to thank my supervisors Dr Jessica Fish and Prof Jon Evans. Your knowledge, guidance and input has been invaluable in helping me complete this project. It has been a privilege to have worked alongside you and I want to thank you for your patience and flexibility throughout the project.

I would also like to express my gratitude to my parents, George and Helen, my sister Dimitra my partner Paul and my aunt Hara whose love and support has sustained me, motivated me and encouraged me throughout all my endeavours. I would have not been able to be where I am without their unconditional support love and belief in me.

I would like to thank my friends Niki and Olga as well as the friends that I made throughout my journey on this course. Thank you all for the support encouragement and understanding you have shown these last few years, and of course a special thank you to my second reviewer Raluca for her hard work!

Foreword

The original thesis project aimed to assess the feasibility and acceptability of a novel photographic-portraiture intervention for adults with non-progressive brain injury. However, for reasons unrelated to COVID this project could not proceed and therefore the current project was agreed.

The MRP proposal submitted for marking to the university was developed for the original project, while the current project proposal was not formally submitted for marking. Therefore, both proposals are included in the appendices for reference.

Chapter 1

Psychometric Properties of the Hotel Test in Acquired-Brain Injury:

A Systematic Review

Word Count: 8091

Prepared in accordance with the author requirements for the Clinical Neuropsychologist
https://www.tandfonline.com/action/authorSubmission?show=instructions&journalCode=ntcn20

Abstract

Objective**:** The Hotel Test is a multitasking assessment of executive functioning (EF) developed in response to the need for ecologically valid assessments of EF in Acquired Brain Injury (ABI). The purpose of the review is to systematically analyse the evidence regarding the psychometric properties of the Hotel Test to guide decisions regarding its use in research and clinical practice.

Methods**:** Six electronic databases were systematically searched in September 2021. Risk of bias was assessed using the COSMIN checklist. Study results on each measurement property were collated and synthesised.

Results**:** Eleven studies were identified reporting data on the measurement properties for at least one of the Hotel Test scores. There was low and very-low quality evidence for convergent and ecological validity respectively, and moderate quality evidence for known-group validity. No studies reported data on content, structural, cross-cultural and criterion validity or reliability (internal consistency, test-retest & interrater reliability and measurement error). The quality of the evidence for responsiveness could not be determined.

Conclusion**:** The review provides some support for the use of the Hotel Test in ABI research and clinical practice. However, the tool should be used cautiously due to the lack of strong evidence regarding its psychometric properties. Reasons for the mixed results are discussed and areas for future research are suggested.

Keywords**:** psychometric properties, executive function, systematic review

Introduction

Background

Non-progressive acquired brain Injury (ABI) is the term used to describe injury to the brain either through any form of traumatic brain injury (TBI) or non-TBI related injuries such as stroke, brain tumour or hypoxic-ischaemic episodes (Menon & Bryant, 2019). ABI is a growing public health concern and it has been described as a leading cause of death and disability in the United Kingdom (UK) as an estimated 1.3 million people in the UK are living with the effects of TBI alone with one hospitalisation due to ABI being recorded every 90 seconds (All-Party Parliamentary Group on Acquired Brain Injury, 2018). The data on the economic impact of ABI estimate costs of £15 billion per year, almost 10% of the healthcare budget (All-Party Parliamentary Group on Acquired Brain Injury, 2018).

ABI has been commonly associated with frontal lobe damage and some of the most debilitating outcomes in people with ABI are arguably associated with impairments in executive function (EF; Jovanovski et al., 2012). Although there is no single agreed upon definition, EF refers to cognitive processes responsible for planning, organising, initiating, monitoring and regulating goal-directed behaviour (Gioia & Isquith, 2004). EF is therefore vital in carrying out aspects of most everyday life tasks including maintaining relationships and employment and managing finances (Jovanovski et al., 2012; Mueller & Dollaghan, 2013). Moreover, EF deficits affect an individual's ability to make use of other intact areas of functioning and hinder self-management efforts thus hindering the ability to successfully apply compensatory strategies (Lewis et al., 2011). EF impairment can therefore have a serious impact on independent adaptive functioning and comprehensive assessment of EF following brain injury is paramount for effective rehabilitation plans.

However, several problems have been highlighted regarding the assessment of EF. Although a variety of measures are available, most traditional measures of EF have been criticised for their lack of ability to predict everyday impairment and there is limited agreement regarding their research and clinical utility (Gioia & Isquith, 2004; Jovanovski et al., 2012; Mueller & Dollaghan, 2013).

Assessment of EF

It is important to acknowledge the inherent complexities of EF assessment in ABI. First, ABI refers to a heterogenous group of injuries and therefore the nature of the impairment can differ according to the brain areas and hence the neurological processes affected. Moreover, there is considerable theoretical debate as to the extent to which EF abilities represent separate processes or share a common underlying process. The lack of agreement concerning the nature of EF and the diversity of the ABI population has led to the development of a variety of diverse tests, and poses challenges for clinicians who are called on to make decisions about appropriate assessment and rehabilitation plans for people with ABI (Mueller & Dollaghan, 2013).

Tests such as the Wisconsin Card Sorting Test (Grant & Berg, 1948), the Trail-Making test (Army Individual Test Battery, 1994) and the Stroop Test (Stroop, 1935), are commonly used in clinical practice. However, they have been criticised for their lack of predictive and ecological validity as they have often failed to discriminate between clinical and non-clinical samples (George & Gillbert, 2018; Manchester et al., 2004). More importantly, it has been demonstrated that performance on these tests can often be spared even in the face of severe everyday executive impairment (Burgess et al., 2006; Shallice & Burgess, 1991). Indeed, Burgess et al. (2006) discussed flaws in the rationale behind the prevalent use of such assessments. In perhaps one of the most well-known demonstrations of traditional EF tests' lack of sensitivity, Shallice & Burgess (1991) discussed how three patients with frontal lobe damage performed normally on classic tests of EF, however had clear EF impairment in daily life. Such results have since been replicated in different neurological populations (Torralva et al., 2012).

One proposed explanation is that the structured testing environment that most classic neuropsychological tests of EF require compensates for the effects of executive dysfunction in patients and may provide a poor environment to elicit such deficits. In other words, assessments are usually performed in a standardised, distraction-free environment to maximise performance. Although, this might be important for the purposes of establishing impairment, it provides little information for how the individual might perform in a more complex ill-structured everyday-life environment (Chaytor et al., 2006; Gioia & Isquith 2004). Furthermore, most of these tests were designed for use with neurologically healthy populations to assess distinct, narrow aspects of EF. As a result, they may fail to meaningfully capture the complexity of real-life EF demands (Gioia & Isquith, 2004). As a result, researchers have worked to develop more ecologically valid measures of EF.

Shallice & Burgess (1991) argue that EF assessments that resemble the unpredictability and complexity of real-life situations might offer advantages in predicting everyday EF deficits. They went on to describe two ecologically valid tests, sensitive to dysexecutive impairments. The first was the Multiple Errands Task (MET), a naturalistic test where patients are asked to complete a number of tasks carried out in a real-life setting whilst following a set of rules. The second, the Six Elements Test (SET), is a desktop assessment of multitasking. Common elements of these assessments include the presence of different sub-goals that need to be achieved whilst patients need to balance competing demands and adhere to rules. Indeed, accumulating evidence suggests that ABI patients with EF deficits perform poorly on this kind of assessment, indicating that they might be useful tools for clinical and research purposes and especially where ecological and predictive validity is concerned (Fish, 2008). Following this, several multitasking tests of EF have been developed and adapted for use in different settings and with different populations (Chaytor et al., 2006).

The Hotel Test

The Hotel Test was developed by Manly et al. (2002). It is a multitasking assessment of EF similar to the SET and simulates a real-life scenario whereby participants need to complete a series of tasks one might have to engage in when working in a hotel. The test comprises six tasks to be completed in 15-minutes (Manly et al., 2002). In summary, participants are asked to imagine they work in a hotel and that within the next 15 minutes they need to try each of a series of tasks to get a "feel" for the work. The tasks include compiling bills, sorting coins for a charity collection, looking up telephone numbers, sorting name tags and proofreading a leaflet. In addition, participants are asked to remember to open and close the hotel garage doors by pressing a button at two predefined times.

As with SET, the total time is insufficient to complete all tasks and therefore it requires planning, self-monitoring, cognitive control and task switching. Outcomes of interest are the total number of tasks attempted, time deviation from the optimal time allocation for each task, whether participants remembered to open the doors or not as well as time deviation for opening the garage doors (Manly et al., 2002).

The Hotel Test has been shown to be sensitive to impaired EF in people with ABI (Manly et al., 2002; Roca et al., 2011) as well as other conditions including dementia and psychosis (between group effect size d= 1.88**,1.24**, 1.81** and 1.7** respectively; Torralva et al., 2009; Torralva et al., 2012). It is brief to administer and materials are easily accessible to most clinicians. It could therefore be a valuable clinical tool to assist diagnosis and facilitate appropriate intervention planning in ABI. Despite promising preliminary information regarding the Hotel Test's utility and validity, to date there is no summary of information regarding the psychometric properties of the test for use in people with ABI.

Aims

The aim of the current study is, , to systematically review, critically appraise, compare and summarise the quality of evidence regarding the psychometric properties of the Hotel Test for use in people with non-progressive ABI.

Methods

The review follows the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines for evaluating studies on measurement properties of patient-reported outcome measures (PROMS; Mokkink et al., 2018). Although the methodology was initially designed to evaluate PROMS, it can also be used for clinician-reported or performance-based outcomes. The guidelines were originally developed to evaluate the most appropriate measure of a construct, from a range of possible measures. However, the guidelines can be used to evaluate the psychometric properties of a single measure (Mokkink et al., 2018). The review also follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021).

The study protocol outlining the aims and methods for the review, was registered with the prospective register of systematic reviews (PROSPERO) in October 2021 (CRD42021285411).

Literature search

Initial searches were conducted with the Cochrane Database of Systematic Reviews (CDSR), PROSPERO and the Database of Abstracts of Reviews of Effects (DARE), to ensure no similar reviews had been published. Articles were then identified through systematic search of the following databases: PsychINFO, Medline, Embase, Web of Science, CINAHL and Scopus. The searches were conducted in September 2021.The search strategy was developed in consultation with a librarian and included a combination of the terms "Hotel Test", "executive functioning" and "multitasking". A more detailed description of the search strategy can be found in appendix 1.1.

In addition to these databases, searches were performed through Google Scholar, the British Library e-thesis online service (EThOS) and ProQuest Dissertations & Theses Database using shortened search criteria. Reference list checks of the identified studies were performed to find any further relevant literature. The search was limited to articles published from 2002 onwards as this is when the paper reporting the original version of the Hotel Test was published (Manly et al., 2002).

Eligibility Criteria

The following inclusion criteria were applied:

- Original studies published in peer-reviewed journals and unpublished PhD theses.
- Studies that discuss any variation of the Hotel Test, in an adult population with non-progressive ABI including TBI, stroke, and any other acquired non-progressive ABI.
- Articles that provide psychometric data on the Hotel Test, relevant to at least one of the following: reliability, content validity, convergent validity, construct validity, discriminant validity, known-group validity and responsiveness to change.
- Intervention studies using the Hotel Test as an outcome measure were used as evidence of responsiveness to change.

- Articles published in English.
- Participants' age >18 years

Exclusion criteria:

- Conference abstracts and non-peer-reviewed published evidence.
- Review articles.
- Articles written in any language other than English.

After duplicates were removed articles were initially screened by title and abstract. Articles that seemed relevant to the review were then screened by full text by two reviewers and disagreements were resolved through discussion. Interrater agreement on study selection was k=.8.

Data extraction

Data was extracted by the first author. A data extraction form was developed that included information on author, year of publication, country, study design, sample size, descriptive information for the patient and control groups where applicable (number of participants, health condition, age and gender), and intervention where applicable (type of treatment). Moreover, the measurement properties assessed by each study and comparator measure used (where applicable) were also recorded.

Risk of Bias

Following data extraction, the studies were assessed for risk of bias using the COSMIN risk of bias checklist (Mokkink et al., 2018; Terwee et al., 2018). Due to time constrains a second reviewer independently rated 60% of the selected studies. Agreement was k=.85 and any disagreements were resolved through discussion. Although available it was not necessary for a third reviewer to be consulted. The COSMIN risk of bias checklist comprises 10 items assessing the methodology for different measurement properties (see table 1 for a description). Items are rated as “very good”, “adequate”, “doubtful” or “inadequate” based on the COSMIN criteria and a “worst score counts” approach is used to determine the overall quality of the methods on a measurement property. For example, if all but one items for reliability in a particular study are assessed as “very good” and the remaining one is assessed as “doubtful” then the overall rating for how well the study assessed reliability is rated as “doubtful”. For studies assessing more than one measurement property all relevant boxes are used independently.

Moreover, for the convergent and ecological validity categories, if a study used more than one comparator measure, box 9 was completed separately for each measure. Evidence for validity was then considered separately if different measures within the same study were assessed as having different quality. For example, if a study was assessed as having “very good” statistical analysis and two comparator measures were assessed as “very good” and “doubtful”, we then considered the study as two separate studies and by applying the worse score counts approach a score of “very good” and “doubtful” overall quality respectively was applied.

Table 1. 1
COSMIN Categories and corresponding Boxes

PROM development / Content validity	1, 2
Structural validity	3
Internal Consistency	4
Cross-cultural validity	5
Reliability	6
Measurement error	7
Criterion validity	8
Construct validity[a]	9
Responsiveness	10

Note. [a] Including convergent, ecological and known-group validity
Adapted from Mokkink et al., 2018

Evaluation of measurement properties

The results of each study on a measurement property were then rated by the first author as sufficient (+), insufficient (–), or indeterminate (?) according to COSMIN's criteria for good measurement properties (Mokkink et al., 2018; Terwee et al., 2018). According to the guidelines, predetermined hypotheses regarding the magnitude and direction of the expected results were developed by the review team if these had not been defined by the study authors. The review team defined hypotheses regarding the expected results for convergent, ecological and known-group validity as well as responsiveness to change.

For convergent and ecological validity, the hypotheses were based on the magnitude and direction of the correlation of the Hotel Test outcome variables and the comparator measures reported in the studies. According to COSMIN guidelines we anticipated a Pearson $r \geq 0.3$ with instruments measuring executive functioning constructs such as multitasking and prospective memory tests. One study reported an association between the Hotel Test and the Culture Fair test (Cattell & Cattell, 1973) which is a general intelligence test. We therefore hypothesised that although a positive correlation would be expected, this would be ≤ 0.30 as the constructs of executive functioning and general intelligence are related but dissimilar (Friedman & Miyake, 2017).

The hypotheses for known-group validity were defined as follows: Differences in the Hotel Test time deviation score and number of tasks score between ABI groups and healthy participants will have a large effect size (Cohen $d \geq 0.8$).

Finally for responsiveness, the available evidence allowed only for calculations of within-group score differences after a goal management training intervention. A recent meta-analysis into the effect of GMT on executive functioning suggested small effect sizes (Stamenova, & Levine, 2018). Therefore, the hypothesis was defined as evidence of within-group score changes for Hotel Test time deviation and number of tasks, with at least a small effect size

(Cohen d ≥ 0.2). As per COSMIN guidelines, hypotheses for construct validity and responsiveness were based on the direction and magnitude of the relationship rather than significance of the results (Mokkink et al., 2018). Where effect sizes were not available, these were calculated using reported means, standard deviations and sample sizes.

Synthesis

One reviewer qualitatively summarised the nature of the existing evidence on measurement properties of the Hotel Test. The results on each measurement property were rated as sufficient (+), insufficient (–), inconsistent (±), or indeterminate (?) if at least 75% of the total studies' results were rated as such for that particular measurement property. For example, evidence for known-group validity was considered sufficient if at least 75% of studies fulfilled the hypothesis for known-group validity.

Finally, one reviewer graded the quality of the summarised evidence according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach as described by COSMIN (Monnik et al., 2018). According to GRADE recommendations the overall quality of the evidence as it pertains to each measurement property is rated as high, moderate, low or very low depending on the level of the overall risk of bias, inconsistency of the results, imprecision (related to the overall sample size) and indirectness.

Results

Study Selection

The search identified a total of 219 records. After duplicates were removed 129 studies were screened for eligibility and 25 were accessed for further screening. Finally, 11 studies were included in the review. The study selection process is illustrated in Figure 1. One study (Dey, 2019) was excluded because quantitative information for the Hotel Test scores was missing as the study combined the scores with other EF scores to yield one total EF score. Moreover, two of the included studies (Roca et al., 2010; Roca et al., 2011) included the same control group, however as the studies included a different ABI group they were considered as two separate studies.

Study Characteristics

None of the studies were aimed specifically at evaluating the measurement properties of the Hotel Test. Three of the included studies were cross-sectional studies looking at associations between theoretically relevant aspects of EF (Banks et al., 2016: Roca et al., 2010, Roca et al., 2011), two were studies validating another test of multitasking or prospective memory (Baylan, 2014; Cullen et al., 2016), three were RCTs of cognitive rehabilitation interventions (Gracey et al., 2017; Levine et al., 2011; Tornas et al., 2016), and three were experiments focussed on theoretical aspects of neuropsychology (Fish, 2008; Fish et al., 2007; Manly et al., 2002). None of the studies reported data on content, structural, cross-cultural and criterion validity, internal consistency, reliability or measurement error. Results are therefore presented only for ecological, convergent and known-group validity and responsiveness to change.

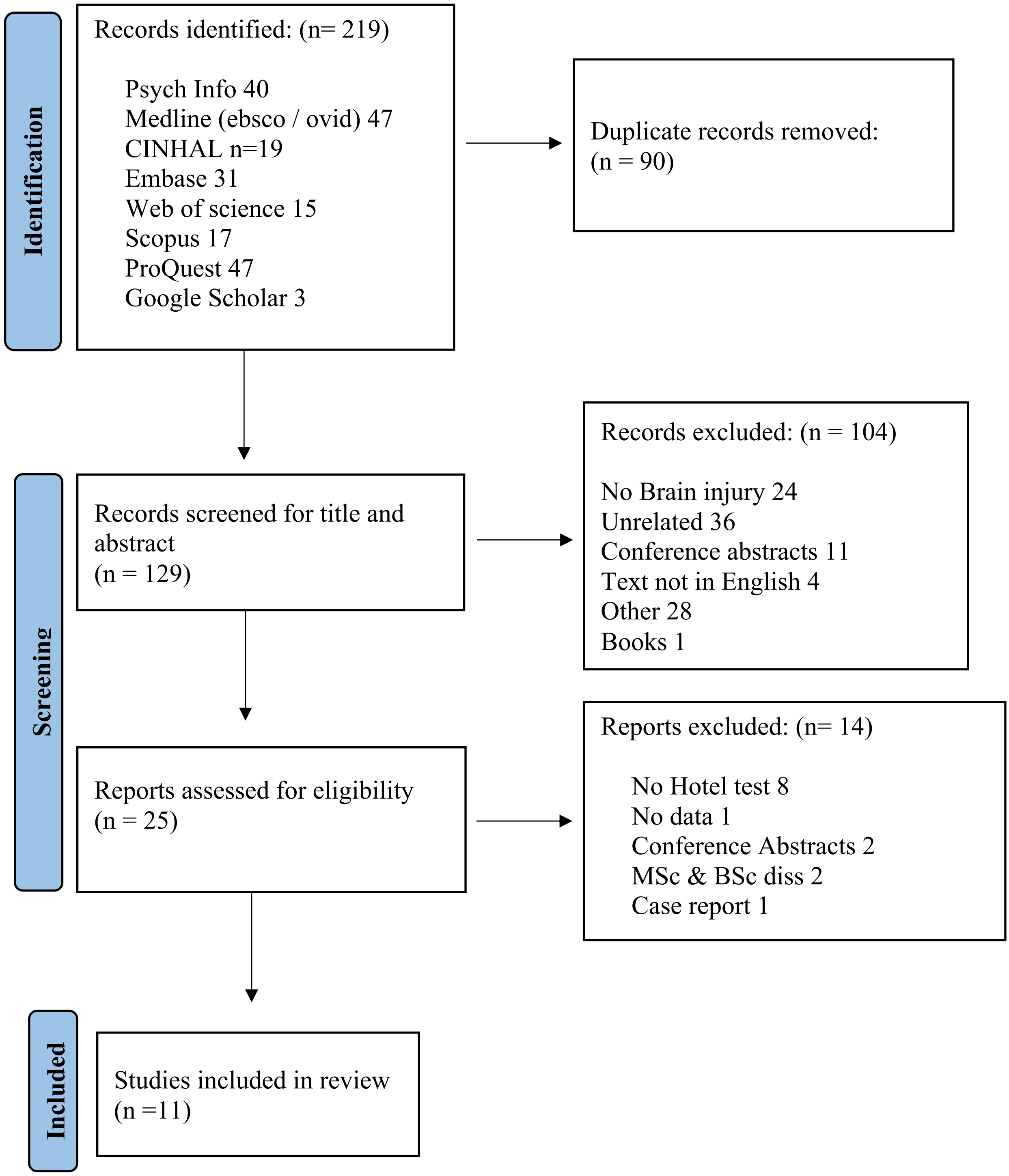

Figure 1.1 PRISMA diagram of the search process

Five studies reported scores on both number of tasks attempted and time deviation. Three of the studies only reported outcomes for number of tasks, two for time deviation and for one of the studies (Gracey et al., 2017), it was unclear which score was reported. Therefore, results are presented separately for the two scores. Details of the included studies can be found in Table 2. Moreover, only one study (Manly et al., 2002) reported results on garage door button pressing and garage door time deviation and therefore sufficient evidence could not be summarised for these outcomes. Finally, it is worth noting that two studies reported slight variations of the Hotel Test. In her experimental study, Fish (2008) removed the prospective memory component (garage door instruction), whilst Baylan (2014) reported outcomes on a modified version of the Hotel Test whereby instead of having to press a button to open the garage doors participants listened to a continuous recording of car registration numbers and were instructed to ring a bell every time a car with a plate ending in CBC arrived.

The evidence per measurement property is detailed below and a summary can be found in Table 3. Further information regarding the details of the methodological appraisal, study results and criteria ratings can be found in Appendix 1.2 and 1.3.

Validity: Convergent

Five studies reported data on convergent validity (Baylan, 2014; Cullen et al., 2016; Fish 2008; Fish et al., 2007; Roca et al., 2010). Two of those studies (Baylan, 2014; Cullen et al.) were rated twice as they used more than one comparator measure that were assessed as having different quality levels. Therefore, seven comparisons are reported. Of those six studies, three were assessed as "very good" (Baylan, 2014; Cullen et al., 2016; Roca et al., 2010), two were assessed as adequate (Cullen et al., 2016; Fish et al., 2007) whilst two were rated as of "doubtful quality" (Baylan, 2014; Fish 2008).

Three studies reported five comparisons for the Hotel Test, number of tasks. In total, four out of five (80%) hypotheses were confirmed, rendering the results "sufficient". The one study (Baylan et al., 2014) reporting a small non-significant correlation (Spearman r= .21) utilised a non-standardised computerised prospective memory task. The pooled results were downgraded to "Low" due to inconsistency and imprecision.

Regarding the Hotel Test time deviation score, five out of eleven reported hypotheses (45%) were confirmed and therefore the results were rated as inconsistent. It is worth noting that one of the studies (Roca et al., 2011) reported a correlation which included both individual with ABI and healthy controls. Although this is likely to have affected the relationship between the reported measures, given the limited evidence available and because the ABI population comprised at least 45% of the sample, it was decided to include the hypothesis in the analysis. Upon further examination of the results no apparent reasons could be determined for the inconsistency and therefore, according to the COSMIN guidance (Mokkink et al., 2018) the overall quality of evidence could not be determined.

Validity: Ecological

Two studies reported correlations between the Hotel Test and measures of everyday prospective memory problems (Baylan, 2014) and a real-life prospective memory task (Fish et al., 2007), reporting a total of 5 hypotheses. One of the studies (Fish et al., 2007) was rated as having doubtful quality whilst box 9 of the COSMIN methodology was used twice for Baylan's (2014) study. The study was rated as "very good" when using a well validated comparator measure however, it was subsequently rated a second time as "inadequate" when reporting data on a non-validated goal management difficulties questionnaire.

One out of three hypotheses (33%) were confirmed for the Hotel Test, number of tasks therefore the results were rated as inconsistent, and the overall quality could not be determined.

Similarly, only one study (Baylan, 2014) reported outcomes on the Hotel Test, Time deviation score. One medium corelation (spearman r= .35) was observed between the Hotel Test, time deviation and a measure of everyday prospective memory problems, whilst only a small correlation was observed for a goal-management measure. As the goal management measure has not been validated and thus was deemed of lower quality, the results were rated as sufficient although they came from the same population. However, the overall quality rating was determined as very low as it was downgraded due to serious inconsistency and imprecision.

Validity: Known group

Six studies were identified that reported data on known-group validity, of which one (Manly et al., 2002) had very good methodological quality, two (Roca et al., 2010; Roca et al., 2011) adequate quality and the remaining three doubtful (Banks et al., 2016; Baylan 2014; Fish, 2008). Four of the studies (Baylan, 2014; Fish, 2008; Manly et al., 2002; Roca et al., 2011) included information on the Hotel Test number of tasks score for known-group validity. Only two of four hypotheses were confirmed and therefore the results were rated as inconsistent, and the overall quality could not be determined.

In addition, all the studies reported data on the Hotel Test time deviation score and of the six hypotheses five (83%) were confirmed. Thus, the pooled results were rated as sufficient for known-group validity/time deviation score. The overall quality was downgraded to moderate for inconsistency as not all results confirmed the predefined hypothesis.

Responsiveness

Three studies were identified from which data on responsiveness to change could be extracted. One study was rated as having Inadequate quality (Gracey et al., 2017) and the remaining two were rated as Doubtful (Levine et al., 2011) and Very good (Tornas et al., 2016). One study

(Gracey et al., 2017), reported no differences in the Hotel Test scores following a brief goal management training programme, however it was unclear which Hotel Test score was reported in the study. The remaining two studies reported outcomes both on the Hotel Test time deviation and number of tasks scores.

One out of two hypotheses was confirmed both for the Hotel Test, number of tasks and time deviation between the two remaining studies. Examination of the type of intervention did not explain the inconsistency as both studies used a goal management training intervention. The overall evidence for responsiveness was therefore rated as inconsistent. Consequently, the overall quality of the evidence could also not be determined.

Table 1.2
Study Characteristics

Author/year	Country	Design	Reported Scores	Measurement properties	Clinical Sample	Control Group
					Condition: n; Age, Mean SD; Gender (n, %)	Condition: n; Age, Mean SD; Gender (n, %)
Banks et al., 2016	USA	Prospective case-control	Time dev	Val: K-gp	mTBI: n=13, 39.3y (14.0); M=9 (69%) F=4 (31%)	HC: n= 11, 37.6y (13.3); M=7 (64%) F= 4 (36%)
Baylan, 2014	UK	Cross-sectional	No tasks, Time dev	Val: Conv, Ecol, K-gp	ABI: n=39; 47.1y (10.3); M= 32 (82%) F= 7 (18%)	HC: n= 16, 40.2y (13.7); M=8 (50%) F= 8 (50%)
Cullen et al., 2016	UK	Cross-Sectional	No tasks	Val: Conv	ABI: n=17; 34.1y (10.5); M=14 (78%) F= 4 (22%)	n/a
Fish, 2008	UK	Cross-sectional	No tasks, Time dev	Val: Conv, K-gp	ABI: n=16; 44.9	HC: n=12, 48.17
Fish et al., 2007	UK	Single cohort	No tasks, Time dev	Val: Conv	ABI: n=20; 40.8y (12.6), M= 15 (75%) F= 5 (25%)	n/a
Gracey et al., 2017	UK	RCT	n/a	Resp	ABI: n= 29; 47.79y (14.72), M= 21 (72%) F=5 (28%)	ABI: n=30; 49.76y (12.94); M=21 (70%) F=9 (30%)
Levine et al., 2011	Canada	RCT	No tasks, Time dev	Resp	ABI: n=11; 48.91y (12.83); M= 8 (73%) F= 3 (27%)	ABI: n=8; 49.25y (13.85); M= 6 (75%) F= 2 (25%)
Manly et al., 2002	UK	Cross-sectional	No tasks; Time dev; GD, GD Time dev	Val: K-gp	TBI: n=10; 32.1y (11.1); M=9 (90%) F=1 (10%)	HC: n=24; 29.29y (8.90); M= 18 (75%) F=6 (25%)
Roca et al., 2010	Argentina	Cross-sectional	Time dev	Val: Conv, K-gp	ABI (tumour): n=20, 55.7y (14.2)	HC: n=25, 55.0 7 (14.4)

Roca et al., 2011	Argentina	Cross-sectional	Time dev	Val: K-gp	ABI BA lesions: n=7; 50y	HC: n= 25 matched for age
Tornas et al., 2016	Norway	RCT	No tasks, Time dev	Resp	ABI: n=33; 42.12y (13.72); M= 19 (58%) F= 14 (42%)	ABI: n=37; 43.57y (12.39); M= 19 (51%) F= 18 (49%)

Abbreviations: No= number, n= number of participants; SD= standard deviation; dev= deviation; Val= Validity; K-gp= Known-group; mTBI= mild Traumatic Brain Injury; M= male; F= female; HC= healthy controls; Ecol= Ecological; AB= Acquired Brain Injury; Conv= Convergent; RCT= Randomised control trial; Resp= Responsiveness; GD= Garage doors.

Table 1.3
Sufficiency of measurement properties

	No of Tasks					Time deviation				
Property	No of Studies	Total Sample n	Results	Rating	GRADE rating	Total Sample n	No of Studies	Results	Rating	GRADE rating
Val: Con	3	77	4/5 (80%) Confirmed	Sufficient (+)	Low (-2)	100	3	5/12 (42%) Confirmed	Inconsistent (±)	Indeterminate
Val: Eco	2	59	1/3 (33%) Confirmed	Inconsistent (±)	Indeterminate	39	1	1/2 (50%) Confirmed	Sufficient (+)	Very Low (-4)
Val: K-gp	4	149	2/4 (50%) Confirmed	Inconsistent (±)	Indeterminate	218	6	5/6 (83%) Confirmed	Sufficient (+)	Moderate (-1)
Resp	3	74	1/3 (33%) Confirmed	Inconsistent (±)	Indeterminate	74	3	1/3 (33%) Confirmed	Inconsistent (±)	Indeterminate

Abbreviations: No= number n= number of participants; GRADE= Grading of Recommendations Assessment Development and Evaluation; Val= Validity; Conv= Convergent; Ecol= Ecological; K-gp= Known-group; Resp= Responsiveness.

Discussion

To the best of our knowledge this is the first systematic evaluation and synthesis of the evidence regarding the psychometric properties of the Hotel Test. We believe that a focus on ABI is particularly important given the prevalence of EF deficits on this population. An investigation of the psychometric properties of the Hotel Test might also be particularly helpful for clinicians given the increasing popularity and interest in naturalistic assessments of EF in clinical practice.

The review identified 11 studies which included data on any of the measurement properties for at least one of the main Hotel Test scores (number of Tasks, Time deviation) within an ABI population. No studies were found reporting evidence on content validity. According to the COSMIN taxonomy content validity refers to the degree to which the tool is an adequate reflection of the underlying construct it is supposed to measure (Mokkink et al., 2018). It is therefore considered one of the most important properties of a measure (Prinsen at el., 2018). Additionally, there were no studies on structural validity. Further, we found no studies reporting data on cross-cultural and criterion validity or reliability (internal consistency, test-retest & interrater reliability and measurement error) of the tool.

Overall, we found the evidence for the measure's convergent and ecological validity was low and very low respectively, whilst there was moderate evidence for known-group validity. The quality of the evidence for responsiveness could not be determined.

We found sufficient evidence for convergent validity of the number of tasks scores however, the quality of the data was low. This was mainly due to the small sample size and inconsistency in the results. Interestingly, one study with very good quality reported a modest correlation between the Hotel Test and the modified SET, a test of multitasking on which the Hotel Test was based (Cullen et al., 2016). A possible explanation for this might, however, relate to practice effects. There is always a novelty effect associated with EF tests both related to content and format, and repeated administration leads to problems with practice effects (Tornas et al., 2016). In the aforementioned study, three multitasking tests were administered and although different in content it is likely that repeated administration led to familiarity with the format and thus improved performance in the tests. Moreover, the evidence for the correlation of the time deviation score with other neuropsychological tests of EF was mixed. This is not uncommon for tests of EF (Chaytor et al., 2006; Rotenberg et al., 2020). Indeed, as it has been suggested different tests of EF tap into different aspects of EF and multitasking assessments elicit everyday executive deficits which may not be easily demonstrated through single task laboratory assessments (Chaytor et al., 2006; Mueller & Dollaghan, 2013). More evidence regarding the processes involved in performance on the Hotel Test is required to investigate this. Additionally, eight out of eleven hypotheses tested the Time Deviation score were from the same study and therefore from the same participants. Although according to COMIN guidance each comparison is treated as a separate hypothesis, it is likely that sample related factors have affected these results, and this should be kept in mind when interpreting the results. Unfortunately, due to the paucity in research, there was limited evidence upon which to draw conclusions from.

Similarly, there was some very low-quality evidence for the ecological validity of the Hotel Test as it pertains to the time deviation score, although the evidence for the number of tasks score was inconsistent. This was mainly due to the paucity of available research as only two studies were identified utilising mostly non-validated measures of everyday prospective memory problems. Moreover, it is worth noting that according to the existing evidence both Hotel Test scores correlated significantly with a validated measure of daily EF. However, the hypothesis for ecological validity was not supported when the Hotel Test sub scores were compared to either a non-validated measure or an everyday prospective memory task (Baylan

2014; Fish et al., 2007). Despite the low quality there was some indication that the Hotel Test might be useful in predicting everyday executive impairment in people with ABI. This is in accordance with research indicating acceptable ecological validity of the Hotel Test in different disorders (Caletti et al., 2013; Torralva et al., 2012). However, more research utilising specific measures of everyday EF difficulties such as the Dysexecutive Questionnaire (Burgess et al., 1996) is warranted.

The evidence pertaining to the time deviation score known-group validity was stronger with large effect sizes observed for the difference in the scores between people ABI and healthy controls. Interestingly, most of the studies on this property selected participants based on the presence of reported EF impairment compared to just the presence of ABI. It could therefore be argued that this represented a more appropriate sample upon which to explore this property. This supports the argument that naturalistic assessment is appropriate in detecting executive dysfunction (Burgess et al., 2006) and suggests that the Hotel Test might be useful in clinical practice. Regarding the number of tasks score however, the evidence was mixed. It is worth noting that all the identified studies reported a statistically significant difference on the Hotel Test number of tasks score between ABI and healthy participants however, in two of those studies a medium and a small effect size was observed (Baylan 2014; Manly et al., 2002). The choice of an effect size compared to statistical significance was however, deemed more appropriate for assessing known-group validity as it provides a better indicator as to whether a measure discriminates accurately between two groups (Manchester et al., 2004). A possible explanation for the lack of sensitivity for the Hotel Test number of tasks score, is that in both studies ABI patients demonstrated near ceiling performances which indicates that mild executive impairment might not be easily detected by this score. Interestingly, the largest effect size was observed in the study which compared healthy participants with those with damage to Brodmann Area 10, which has been proposed to play a particular role in multitasking (Roca et al., 2011). Moreover, the presence of ceiling effects would suggest that this score might not be appropriate for evaluating the effects of interventions as it would not allow for variation in performance.

There was also mixed evidence for responsiveness for both scores of the Hotel Test. Again, there was paucity in research as only three intervention studies were identified two of which were of doubtful and inadequate quality with regard to evaluating responsiveness. Overall, there are challenges associated with measuring responsiveness especially in EF that are worth mentioning. When it comes to EF interventions it has been reported that although changes might be observed on questionnaires of everyday impairment, studies fail to observe such changes on tests of EF (Tornas et al., 2016). It is likely that treatment effects might therefore be a result of the use of compensatory strategies rather than changes in EF and therefore not evident during neuropsychological tests. In this case one would not expect a change in EF scores. However, more research into the underlying mechanisms of EF intervention is needed. Moreover, in keeping with goal-management intervention literature, in this review all studies demonstrated a trend towards improvement on the Hotel Test both in the intervention and control groups. This again brings up issues of test-retest reliability and raises questions about the use of neuropsychological tests as outcome measures as change may arise from practice effects rather than an improvement in the underlying construct (Tornas et al., 2016). This is problematic for measurement of responsiveness if it is defined as a measure's ability to detect change when it has actually taken place. Indeed, there is considerable debate in the literature regarding the best way to capture responsiveness and many have argued that effect sizes might not be appropriate (Terwee et al., 2003). Instead, a correlational approach whereby the outcome of interest is how performance on one measure changes in relation to performance on similar measures before and after an intervention, has been deemed more appropriate (Terwee et al., 2003). Unfortunately, we could not identify any studies providing such information and therefore furfure studies are needed which would include data on change between multiple measures of EF.

Finally, there are several limitations regarding the literature which affect the conclusions that can be drawn from the results. First, there was limited research available for conclusions to be drawn from. This is especially relevant for applying the COSMIN methodology which specifies that for sufficiency of a measurement property to be established at least 75% of hypothesis should be confirmed (Mokkink et al., 2018). Consequently, with a small number of available results on hypotheses even small variations in the evidence can have a strong impact on the overall conclusion and the higher the degree of agreement between studies needed to reach a 75% consensus. Moreover, the quality of the studies as it relates to the evaluation of the measurement properties of the Hotel Test was in many cases limited. This was not surprising as none of the studies' primary objective was to evaluate the psychometric properties of the tool.

Future Directions

In the future, more studies are needed that specifically aim to examine the psychometric properties of the Hotel Test in order to provide higher quality evidence from which to draw conclusions. More specifically, future research should focus on providing evidence pertaining to the content validity of the Hotel Test and establishing whether the Hotel Test measures one or more constructs. Given the multitude of processes involved in EF assessment, a better understanding of the constructs measured by the Hotel Test is needed which will allow for better informed hypotheses regarding associations with various instruments of EF. Moreover, studies need to provide a clear description of the ABI population and the nature of brain damage as well other clinically important variables such as premorbid intellectual functioning and presence of mental health disorders. Finally, it is important for future research to take into consideration the heterogeneity of cognitive impairment in ABI. Given that ABI does not always lead to EF impairment future research needs to provide more detailed evidence regarding performance of the test based on different types of injury, ability, EF deficit etc.

Strengths and Limitations of the review

The current study utilised a comprehensive and systematic selection process which ensured that most of the relevant available research was included in the study, whilst the application of the COSMIN methodology is a particular strength of the review as it allows for a rigorous evaluation of the evidence. Moreover, the inclusion of gray-literature is another strength as it reduces publication bias. However, there are several limitations that need to be acknowledged. First, the inclusion of articles only written in English meant that literature from other populations might have been missed. In addition, due to feasibility issues authors were not contacted for additional information when this was not reported in the original publication. Most importantly, the evaluation of the measurement properties was conducted by one reviewer, and this should be considered when interpreting the results as it introduces the risk of bias. Finally, a meta-analysis of known-group validity effect sizes would have been useful. However, this was not possible due to time constrains.

Conclusions

The current systematic review has important implications for clinical practice and research. The evidence, albeit limited, points to the usefulness of the Hotel Test for use in clinical practice for diagnostic as well as potentially predictive purposes. However, the quality of the evidence is low and caution is needed in the use and interpretation of the different scores. As discussed, the lack of strong evidence for the psychometric properties of the Hotel Test could be explained by a variety of factors pertaining to challenges associated with research in ABI populations, methodological limitations of the available research and issues with the tool itself. The current systematic review further highlights the need for improved research methods and reporting standards in the assessment of EF deficits in ABI.

References

All-Party Parliamentary Group on Acquired Brain Injury. (2018). Acquired brain injury and neurorehabilitation: time for change. https://cdn.ymaws.com/ukabif.org.uk/resource/resmgr/campaigns/appg-abi_report_time-for-cha.pdf

Army Individual Test Battery. (1944). Manual of directions and scoring. Washington, DC: War Department, Adjutant General's Office.

Banks, S. D., Coronado, R. A., Clemons, L. R., Abraham, C. M., Pruthi, S., Conrad, B. N., Morgan, V. L., Guillamondegui, O. D., & Archer, K. R. (2016). Thalamic functional connectivity in mild traumatic brain injury: longitudinal associations with patient-reported outcomes and neuropsychological tests. *Archives of physical medicine and rehabilitation*, *97*(8), 1254-1261.

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L. A., Dawson, D. R., Anderson, N. D., Gilbert, S. J., Dumontheil, I. & Channon, S. (2006). The case for the development and use of" ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the international neuropsychological society*, *12*(2), 194-209.

Burgess, P.W., Alderman, N., Wilson, B.A., Evans, J.J., & Emslie, H. (1996). The dysexecutive questionnaire. In B.A. Wilson, N. Alderman, P.W. Burgess, H. Emslie, J.J.

Evans. (1996). *Behavioural Assessment of the Dysexecutive Syndrome (BADS)*. London: Pearson Assessment.

Baylan, S. M. (2014). Imaging the effects of cognitive rehabilitation interventions: developing paradigms for the assessment and rehabilitation of prospective memory (Doctoral dissertation, University of Glasgow). ProQuest Dissertations and Theses Global.

Caletti, Elisabetta, Riccardo Augusto Paoli, Alessio Fiorentini, Michela Cigliobianco, Elisa Zugno, Marta Serati, Giulia Orsenigo et al. "Neuropsychology, social cognition and global functioning among bipolar, schizophrenic patients and healthy controls: preliminary data." *Frontiers in human neuroscience* 7 (2013): 661.

Cattell, R. B., & Cattell, A. K. S. (1973). Measuring intelligence with the culture fair tests. Champaign, IL: Institute for Personality and Ability Testing.

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of clinical neuropsychology*, *21*(3), 217-227.

Cullen, B., Brennan, D., Manly, T., & Evans, J. J. (2016). Towards validation of a new computerised test of goal neglect: Preliminary evidence from clinical and neuroimaging pilot studies. *Plos one*, *11*(1), e0148127.

Dey, A. K. (2019). Functional neural correlates of cognitive impairment and neural compensation in mild to moderate Cerebral Small Vessel Disease (Doctoral dissertation, University of Toronto). ProQuest Dissertations and Theses Global.

Fish, J. E. (2008). Assessment and Rehabilitation of Prospective Memory In Acquired Brain Injury (Doctoral Dissertation, University of Cambridge).

Fish, J., Evans, J. J., Nimmo, M., Martin, E., Kersel, D., Bateman, A., Wilson, B. A., & Manly, T. (2007). Rehabilitation of executive dysfunction following brain injury: “Content-free” cueing improves everyday prospective memory performance. *Neuropsychologia*, *45*(6), 1318-1330.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186-204.

George, M., & Gilbert, S. (2018). Mental Capacity Act (2005) assessments: Why everyone needs to know about the frontal lobe paradox. *The Neuropsychologist*, *5*(1), 59-66.

Gioia, G. A., & Isquith, P. K. (2004). Ecological assessment of executive function in traumatic brain injury. *Developmental neuropsychology*, *25*(1-2), 135-158.

Gracey, F., Fish, J. E., Greenfield, E., Bateman, A., Malley, D., Hardy, G., Ingham, I., Evans, J. J. & Manly, T. (2017). A randomized controlled trial of assisted intention monitoring for the rehabilitation of executive impairments following acquired brain injury. *Neurorehabilitation and neural repair*, *31*(4), 323-333.

Grant, D.A., & Berg, E.A. (1948). A behavioural analysis of degree or reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology, 38*, 404–411.

Jovanovski, D., Zakzanis, K., Ruttan, L., Campbell, Z., Erb, S., & Nussbaum, D. (2012). Ecologically valid assessment of executive dysfunction using a novel virtual reality task in patients with acquired brain injury. *Applied Neuropsychology: Adult*, *19*(3), 207-220.

Levine, B., Schweizer, T. A., O'Connor, C., Turner, G., Gillingham, S., Stuss, D. T., Manly, T., & Robertson, I. H. (2011). Rehabilitation of executive functioning in patients with frontal lobe brain damage with goal management training. *Frontiers in human neuroscience*, *5*, 9.

Lewis, M. W., Babbage, D. R., & Leathem, J. M. (2011). Assessing executive performance during cognitive rehabilitation. *Neuropsychological rehabilitation*, *21*(2), 145-163

Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain injury*, *18*(11), 1067-1081.

Manly, T., Hawkins, K., Evans, J., Woldt, K., & Robertson, I. H. (2002). Rehabilitation of executive function: facilitation of effective goal management on complex tasks using periodic auditory alerts. *Neuropsychologia*, *40*(3), 271-281.

Menon, D. K., & Bryant, C. (2019). Time for change in acquired brain injury. *The Lancet Neurology*, 18(1), 28.

Mokkink, L. B., De Vet, H. C., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*(5), 1171-1179.

Mueller, J. A., & Dollaghan, C. (2013). A systematic review of assessments for identifying executive function impairment in adults with acquired brain injury. *Journal of Speech, Language, and Hearing Research*, 56, 1051–1064.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E. and Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M. Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*, *372*.

Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*(5), 1147-1157.

Roca, M., Parr, A., Thompson, R., Woolgar, A., Torralva, T., Antoun, N., Manes, F., & Duncan, J. (2010). Executive function and fluid intelligence after frontal lobe lesions. *Brain*, *133*(1), 234-247.

Roca, M., Torralva, T., Gleichgerrcht, E., Woolgar, A., Thompson, R., Duncan, J., & Manes, F. (2011). The role of Area 10 (BA10) in human multitasking and in social cognition: a lesion study. *Neuropsychologia*, *49*(13), 3525-3531.

Rotenberg, S., Ruthralingam, M., Hnatiw, B., Neufeld, K., Yuzwa, K. E., Arbel, I., & Dawson, D. R. (2020). Measurement properties of the multiple errands test: A systematic review. *Archives of Physical Medicine and Rehabilitation*, *101*(9), 1628-1642.

Shallice, T., & Burgess, P. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain 114*, 727–741.

Stamenova, V., & Levine, B. (2019). Effectiveness of goal management training® in improving executive functions: A meta-analysis. *Neuropsychological rehabilitation*, 29(10), 1569–1599.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Terwee, C. B., Dekker, F. W., Wiersinga, W. M., Prummel, M. F., & Bossuyt, P. M. M. (2003). On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Quality of life research*, *12*(4), 349-362.

Terwee, C. B., Prinsen, C. A., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research*, *27*(5), 1159-1170.

Tornås, S., Løvstad, M., Solbakk, A. K., Evans, J., Endestad, T., Hol, P. K., Schanke, A. K., & Stubberud, J. (2016). Rehabilitation of executive functions in patients with chronic acquired brain injury with goal management training, external cuing, and emotional regulation: a randomized controlled trial. *Journal of the International Neuropsychological Society*, *22*(4), 436-452.

Torralva, T., Roca, M., Gleichgerrcht, E., Bekinschtein, T., & Manes, F. (2009). A neuropsychological battery to detect specific executive and social cognitive impairments in early frontotemporal dementia. *Brain*, *132*(5), 1299-1309.

Torralva, T., Strejilevich, S., Gleichgerrcht, E., Roca, M., Martino, D., Cetkovich, M., & Manes, F. (2012). Deficits in tasks of executive functioning that mimic real-life scenarios in bipolar disorder. *Bipolar Disorders*, *14*(1), 118-125.

Chapter 2

The Hotel Test:

Normative data on a measure of multitasking from the UK adult population

Word Count: 8993

Prepared in accordance with the author requirements for the Journal of Clinical and Experimental Neuropsychology, https://www.tandfonline.com/action/authorSubmission?show=instructions&journalCode=ncen20

Plain language summary

Title: The Hotel Test: Normative data on a measure of multitasking from the UK adult population

Background: Executive functioning (EF) describes a set of skills used to plan and organise action, pay attention and manage time. Multitasking has been identified as a very efficient way to assess EF skills**.** Assessment of executive functioning is important for clinicians to be able to understand the impact that EF impairment will have on an individual's life. However, most traditionally used tests fail to accurately predict everyday-life difficulties. The Hotel Test is a multitasking test that may identify everyday EF deficits more accurately than existing tests. It can be easily assembled at minimal cost and could be valuable in clinical practice. However, so far it has mainly been used in small studies, and if it is to be used in clinical practice, we need to know how people without known brain/health conditions perform on it.

Aims: The study examined Hotel Test performance in healthy adults and analysed if multitasking performance varied according to a person's age, sex or intellectual ability. We also investigated some of the technical features of the test and more specifically how performance on the Hotel Test relates to performance with other neuropsychological measures.

Methods: The study used data from 648 people who completed the Hotel Test alongside other cognitive tests as part of a study on cognition and ageing. A series of regression analyses were used to produce normative data and to investigate if multitasking performance was influenced by either demographic factors or performance on other cognitive measures.

Main Findings and Conclusion: Norms for the Hotel Test were produced, and a calculator is included for clinicians to use. Age, education and IQ had an impact on how well people perform on the Hotel Test. Additionally, multitasking performance had a small correlation with performance on other neuropsychological measures. The study provides information for clinicians who are interested in using the Hotel Test in research and clinical practice and will increase the clinical utility of the test.

Abstract

Background: There is a need for ecologically valid tests of multitasking in neuropsychological assessment and rehabilitation. The Hotel Test is an ecologically valid multitasking test of executive functioning (EF) that has shown promise in identifying EF deficits and could be a valuable clinical tool. However, no normative data is available for the test, limiting its utility.

Aims: We aim to improve the utility of the Hotel Test by producing normative data and reporting on its psychometric properties.

Methods: The study utilises existing data from a study on cognition and ageing to produce regression-based norms for the UK population, controlling for the effects of age, education and IQ. Moreover, validity is assessed by analysing the relationship between performance on the Hotel Test and other cognitive measures**.**

Results: Age, education and IQ were significant predictors of multitasking performance. Regression-based norms were developed and calculator and instructions for a companion programme are provided. Small correlations were identified demonstrating divergent validity between Hotel Test performance and relevant neuropsychological measures.

Conclusions: The utility of the test has been enhanced by developing comprehensive norm. The broader results enable clinicians and researchers to make better informed decisions regarding its use.

Keywords: Hotel Test; normative data; regression-based norms; executive functioning

Introduction

Executive function (EF) is a multidimensional construct used to describe the set of control processes required for organising, guiding and monitoring goal-directed behaviour (Gioia & Isquith, 2004; Baggetta & Alexander, 2016). These functions develop in childhood, increase in adolescence, decrease with ageing (Banich, 2009), and are seen as crucial in cognitive academic and social development and independent living (Baggetta & Alexander, 2016).

Executive dysfunction is evident in a number of conditions associated with disruption to frontal lobe function including brain injury, dementia, attention-deficit-hyperactivity disorder (ADHD) and various psychiatric illnesses (Roca et al., 2013; Torralva et al., 2013; Gracey at al., 2017). EF Impairments can severely disrupt the individual's ability to initiate, terminate or shift behaviour, maintain attention, self-monitor, and learn new tasks and therefore can have a detrimental impact on everyday life ability to engage in work or education, live independently and sustain relationships (Mole & Demeyere, 2020).

Given the multitude of ways executive processes are involved in everyday functioning, accurate assessment is paramount not only for contributing to an accurate diagnosis but also for predicting and evaluating the effects of EF deficits in everyday life and subsequently planning interventions.

Classic Assessment of EF

Several tests have been developed to assess aspects of EF including the Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948), the Stroop Test (Stroop, 1935), the Tower of London (ToL; Shallice, 1982) and the Trail-Making test (TMT; Army Individual Test Battery, 1994) among many others. However, these tests have often been criticised for their limited ability to predict real life impairment (Shallice & Burgess, 1991). Of particular interest has been the "frontal lobe paradox" whereby patients with damage to frontal brain areas perform within normal limits on classic tests of executive functioning, despite demonstrating profound impairment in everyday life situations that require EF (Burgess et al., 2006). Indeed, Chaytor et al. (2006) demonstrated that the WCST, the TMT and the Stroop tests only accounted for 18-20% of the variance in everyday EF ability. Moreover, after reviewing the literature, they noted that in studies of ecological validity of classic EF tests there was either no significant association or where there was a significant association, variance explained ranged from 27-65%. They concluded that a large proportion of the variance of everyday EF skills is still unaccounted for (Chaytor et al., 2006).

Classic EF tests tend to tap into narrow components of EF and have been criticised for not being able to meaningfully represent the complexity of real life (Burgess et al., 2006; Chan et al., 2008; Gioia & Isquith, 2004). It has also been argued that the structure inherent to most neuropsychological tests, including taking place in an environment free from distraction and having clear instructions for what is required, and directions for when to start and stop, effectively compensates for any deficits in EF (Shallice & Burgess, 1991).

Ecological Validity of EF assessment

Ecological validity in the neuropsychological context is defined as the relationship between an individual's performance on a neuropsychological test and their behaviour in different real-life settings (Sbordone, 1996 as cited in Lamberts at al., 2010). Inherent in the definition of ecological validity in neuropsychological tests are the concepts of verisimilitude and veridicality. Verisimilitude refers to the representativeness of the tests and its ability to approximate a real-life task, while veridicality refers to the generalisability or in other words the degree to which performance on the test predicts day-to-day functioning (Gioia & Isquith, 2004).

Shallice and Burgess (1991) argued that EF abilities such as the ability to plan and multi-task are fundamental in everyday activities and heavily affected by frontal lobe damage and therefore, everyday-life EF deficits will become more apparent through multitasking tasks with limited structure and feedback from the clinician. Additionally, the shift in clinical assessment from identifying specific deficits in attempt to localise brain damage towards estimating the impact on functioning and the needs for the intervention has led to a shift in considerations of test validity with clinicians and researchers calling for the development of more ecologically valid assessments, particularly in the domain of EF (Burgess et al., 2006; Chan et al., 2008; Gioia & Isquith, 2004). Therefore, in recent years researchers have developed EF tests that mimic multitasking situations that could be encountered in everyday life (Burgess et al., 2006; Manly et al., 2002).

Multitasking tests

Shallice & Burgess (1991) created the Multiple Errands Task (MET) whereby patients are asked to complete several tasks within a shopping precinct whilst following a number of rules. The MET is considered to have high verisimilitude as it represents an everyday situation in a real-life setting whilst it has been shown to possess high veridicality (Burgess et al., 2006). However, despite the MET's clinical usefulness, it is often not feasible due to the complexity of its implementation while computerised and recently developed virtual reality versions such as the Jansari assessment of Executive Function, require technical equipment and expertise which is often not readily available to clinicians (Jansari et al., 2014; Nalder et al., 2017).

Other ecologically valid multitasking tests of EF are the Six Elements tests of the Behavioural Assessment of the Dysexecutive Syndrome (SET BADS, Shallice & Burgess, 1991) and the Executive Secretarial Task (EST, Lamberts et al., 2010). The SET is a tabletop test where the patient is asked to attempt at least some of six basic tasks within a fixed time-period. As insufficient time is given to complete the tasks emphasis is given in the patient's ability to keep track of time and switch activities (Shallice & Burgess, 1991). However, the SET lacks verisimilitude and more crucially is difficult to meaningfully interpret unless the complete BADS battery has been administered. Moreover, the EST; a three-hour task simulating a job assessment, has been designed with verisimilitude in mind and has been shown to be sensitive to EF deficits and to have both concurrent and ecological validity (Lamberts et al., 2010). However, it is time consuming and therefore potentially difficult to implement in clinical practice.

The Hotel Test

Manly et al., (2002), developed a 15-minute multitasking assessment simulating a hotel work environment. The test is based on the SET of the BADS but incorporates a simulation of real-life tasks in a hypothetical real-life environment. The test was later modified to include 5 tasks that need to be completed within 10 minutes (Fish & Manly, personal communication). As with the SET, time is insufficient to complete all tasks and therefore it requires planning, self-monitoring, cognitive control and task switching (Manly et al., 2002). Measures derived include the number of tasks attempted as well as time deviation from the optimal time allocation for each task (Manly et al., 2002). Although the Hotel Test only approximates a real-life situation, and therefore might possess less verisimilitude than tasks such as the EST, it is less constrained than most popular neuropsychological assessments (Manly et al., 2002).

The Hotel Test has been shown to have high sensitivity in assessing EF deficits in various clinical conditions such as brain injury (d=1.88; Manly et al., 2002), frontotemporal dementia (d= 1.81; Torralva et al., 2009), attention deficit-hyperactivity disorder (Torralva et al., 2013), and bipolar disorder (d=1.7; Torralva et al., 2012) with classic test of EF including the WCST failing to differentiate between high functioning patients and controls (Torralva et al., 2009).

Furthermore, one group of studies has demonstrated that differences on EF as measured by classic EF tests may be largely explained by reductions in “general intelligence” (g) in patients with frontotemporal dementia, frontal focal lesions and Parkinson’s disease compared to healthy controls. However, differences in the Hotel Test performance persist even after the influence of g is controlled for indicating that this type of ecological multitasking assessment taps into more unique aspects of EF (Roca et al., 2013; Roca et al., 2012; Roca et al., 2010;).

The Hotel Test is brief to administer and materials are easily accessible to most clinicians. It could therefore be a valuable clinical tool to assist diagnosis and facilitate appropriate intervention planning. However, no normative data is available for the Hotel Test thus limiting its clinical and research utility.

Aims

The primary aim of the current study is to produce normative data for the Hotel Test. The study also aims to examine the association between demographic factors and Hotel Test performance and include significant predictors in the produced norms.

Moreover, the secondary aim is to assess the validity of the test by exploring the correlations between Hotel Test scores and performance on tests of other core cognitive functions including general cognitive ability, fluid reasoning, attention, speed of processing, memory, and another measure of EF.

Methods

Design

The study is a retrospective data analysis based on data collected from the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) repository (available at http://www.mrc-cbu.cam.ac.uk/datasets/camcan/, Shafto et al., 2014; Taylor at al., 2017). The original Cam-CAN study is a large-scale project aiming to provide data on the neural underpinnings of cognitive ageing (Taylor et al., 2017). The study was conducted in three stages and contains cognitive, epidemiological and neuroimaging data (see Shafto et al.). For the current study data from stages I and II were obtained.

Power Calculation

Norms

The American Psychological Association (APA, 1999) suggests that for traditional norming procedures the sample should include no less than 100 participants per factor group (e.g., age category). However, it has been demonstrated that for regression-based norming a considerably lower sample is needed (Lenhard & Lenhard, 2020). Lenhard & Lenhard (2020) noted that regression-based norming required a four times smaller sample compared to conventional norming to achieve the same quality. Moreover, Oosterhuis et al. (2016) suggest that for a short test, a sample greater than 100 is adequate to produce reliable regression-based norms.

Validity

Based on the available sample size a post hoc calculation using G*Power indicated that the study would have 100% power to detect a correlation of .6 or greater, between the Hotel Test and other neuropsychological measures (Faul et al., 2007).

Participants

Participants were adults from the general population. The overall sample consisted of 648 adults, resident in the UK in the Cambridge City area, and recruited via primary care lists (for sampling procedures see Shafto et al., 2014).

Inclusion Criteria

For a detailed description of the Cam-CAN study inclusion criteria and sampling methods see Shafto et al. (2014). For the normative study specific inclusion criteria were:

- Age: 18-90
- No history of neurological or major psychiatric disorder
- Fluent in English
- No current symptoms of depression or anxiety
- No current alcohol and substance abuse problems

Measures

Background information

Background demographic data was obtained on age, sex, education, ethnicity, and English language history. Moreover, data was requested on the Hospital Anxiety and Depression Scale (HADS, Zigmond & Snaith, 1983). The HADS is a 14-item self-report scale which assesses symptoms of overall psychological distress as well as depression (7-items) and anxiety (7-items). Data from several measures assessing cognitive ability as well as functions related to those assessed by the Hotel Test was obtained in order to assess the test's validity as indicated by significant relationships between scores on the Hotel Test and scores on related neuropsychological measures.

Primary outcomes

The Hotel Test (Manly et al., 2002)

The Hotel Test is an EF test investigating skills that are required for planning and multitasking. The original test comprised five tasks that participants might be required to complete if they were working in a hotel and required 15 minutes to be administered. In the Cam-CAN study a modified 10-minute version of the test was used (Fish & Manly, personal communication; see also Shafto et al., 2014). Participants are asked to imagine that they work in a hotel in which their manager asks them to try each of five tasks in the next 10 minutes to get a feel of what it might be like working there. Participants are asked to spend as much time as possible on each task. As the tasks cannot be completed in the time given participants must divide their time to ensure they engage with every task. The tasks include compiling bills, sorting money from a charity collection, sorting playing cards, sorting conference labels by name and proofreading the hotel leaflet. All materials are laid out at the table and in view of the participant. A clock is available and participants are allowed to check the time as often as they require. If the participant does not switch task after the first five minutes, they are remined that the aim is to try all five tasks. The outcomes of interest for the Hotel Test are number of tasks attempted, with a score of five indicating optimal performance, and time allocation on each task. Optimal time allocation would be two minutes on each task. Therefore, the total amount of extra/less time people spend on each task from the optimal time of two minutes is added and the total time deviation from optimal time allocation score is calculated. A total time deviation score of zero would therefore indicate perfect time allocation with higher scores indicating weaker performance.

Spot the Word (STW, Baddeley et al., 1993)

The test assesses familiarity with very low-frequency vocabulary words and provides an estimate of optimal/premorbid cognitive ability. It involves presenting pairs of items comprising one word and one non-word, with the individual required to point to the real word. The STW raw scores were transformed into scaled scores, using the original test norms.

Cattell Culture Fair Test (Cattell, 1971)

This is a pen and paper test of fluid intelligence. Scale 2, Form A of the test was used. It comprises four non-verbal visual puzzle tests yielding one score each which are combined to provide a total score.

Proverb Comprehension (Shafto et al., 2014)

Proverb Comprehension is a task that assesses aspects of EF including abstraction. Participants are asked to explain the meaning of three common proverbs and answers are scored 0, 1 and 2 for an incorrect, partially correct and fully correct answer respectively.

The Wechsler Memory Scale Third UK edition (WMS-III UK), Logical memory subtests Wechsler, 1999)

The logical memory scale of the WMS-III was administered. The test assesses narrative memory and long-term narrative memory under free recall conditions.

Addenbrooke's Cognitive Examination-Revised (ACE-R, Mioshi et al., 2006)

ACE-R is a brief battery used for the screening of cognitive impairment. It provides an evaluation of five cognitive domains (attention/orientation, memory, language, fluency and visuospatial ability). The total ACE-R score was used for this study.

Response time measurements (Shafto et al., 2014)

A “simple” response time task (SRT) and a “choice” response time task (CRT) were included as a measure of processing speed and inhibition. In SRT, people looked at a picture of a hand with blank circles above each finger. When the index finger circle turned black, they pressed with their index finger on a response box as quickly as possible. There are 50 trials.

In the CRT, on each of 67 trials any of the circles could turn black, and the participants needed to press the corresponding finger. The median response time score across all trials was used for the purposes of this study.

Data Analysis

Data screening and analysis was conducted using SPSS.28.0 (IBM, 2021). Descriptive statistics for the demographic and cognitive variables are reported with means and standard deviations (SD) or their non-parametric equivalents (median, interquartile range) for non-normally distributed variables and with frequencies and percentages for categorical variables.

Data was checked for outliers and for normality through skewness and kurtosis and examination of the scatterplots. The Shapiro-Wilk and Kolmogorov Smirnov tests of normality were not considered as it has been suggested that for large samples, they produce a significant result even for very small deviations from normality (Field, 2018).

Three cases were identified with a maximum Hotel time deviation score of 960s and a number of tasks score of 1 and were removed from the dataset. As part of the Hotel Test administration procedure, participants are reminded that the main goal of the task is to attempt part of each of the five tasks. Therefore, if only one task was completed it was possible that the data was either entered in error or that the participant had failed to understand the instructions either because of comprehension difficulties, hearing problems etc., invalidating the score. Moreover, 5 participants were identified with scores suggesting clinical levels of anxiety and 1 of depression as indicated by a HADS score > 15 and were also excluded from the analysis. The final sample consisted of 648 participants (see table 2.1 for details).

All Hotel Test scores as well as ACE-R, RT and Proverbs scores were not normally distributed. A square root transformation was applied for the Hotel Time deviation scores which improved the distribution of the scores. However, no transformation improved any of the other data and therefore regression-based norms were produced only for the Time deviation score.

The Mann–Whitney test was used to investigate associations between sex and Hotel Test performance and Pearson's correlation (r) or the non-parametric equivalent (Spearman's rank) was used to determine associations between the Hotel Test and associated tests.

Norming Procedure

Regression based-norms were calculated using procedures previously published (Rivera & Arango-Lasprilla, 2017; Van Breukelen & Vlaeyen, 2005). First, multiple regression was used to determine the effect of age (in years), sex (coded as 1 for male and 0 for female), education (in years) and premorbid IQ on scores for the Hotel Test (time deviation). All predictors were entered in the model in one block and non-significant predictors were subsequently excluded and the model was re-run. Cases with missing values were excluded listwise. Alpha level was set at 0.05.

Years in education were computed by subtracting 5 years from the age people indicated they left school. We considered 22 years in education as the maximum upper limit. Due to what is known about the relationship between age and EF, age was squared (age^2) to assess for the quadratic effects. Age scores were centred (AgeC= Age – mean age) before quadratic terms were calculated to avoid multi-collinearity (Van Breukelen & Vlaeyen, 2005). Age-centred and age^2 were both added as predictors to the model.

To ensure all regression assumptions were met we examined the normality of residuals through Q–Q plots and non-significant tests of normality of the residuals. Multicollinearity was assessed by Pearson's r correlations < 1 as well as a variance inflation factor (VIF) < 10 and a tolerance value < 1 (Field, 2018; Kutner et al., 2005). The presence of influential cases was assessed by checking for standardised residual values > 3 and Cook's distance > 1 (Field, 2018). No influential cases were detected. Finally, homoscedasticity was determined through visual examination of the Q-Q plots and a non-significant Breusch-Pagan & Koenker test.

After the final regression model was established a 4-step procedure was followed to develop regression-based normative data. First, Predicted Hotel Time deviation scores (Y) were calculated using the final regression equation, based on β weight values for all predictor variables and their predictive constant. Following this, residual values (Re) are calculated by subtracting participants' predicted scores from the actual observed scores. The residuals are then standardised. Standardised Z scores are obtained by dividing the residual error value by the standard deviation of the residual (Crawford et al., 2012). Finally, the percentile values of standardised residuals are calculated using the standard normal cumulative distribution function.

Norm Calculation

In order to enhance clinical utility and to make the norms accessible for clinical use and research purposes an Excel calculator was developed (Appendix 2.1). Clinicians can enter individual Hotel Time deviation raw scores and the calculator will produce a Z score as well as the expected percentile for that score. Moreover, The RegBuild_MR_Raw.exe computer programme (Crawford et al., 2012) creates regression equations using raw data from a normative sample. The programme allows users to compare an individual's predicted score with the actual obtained score using the generated equation. The programme only accepts files with a .txt format and therefore, data from the study has also been prepared in this format for use with programme to enhance clinical utility (Appendix 2.2). The programme is available at: https://homepages.abdn.ac.uk/j.crawford/pages/dept/RegBuild_MR.htm. Clinicians need to use the score conversion Excel file located in Appendix 2.3 which also contains instructions on how to use the programme. Note that the programme only accepts files containing a maximum of 499 cases and therefore some data was randomly deleted. Norms using the excel method and the programme were comparable (see results), indicating that data was omitted at random.

Ethics

The Cam-CAN study was conducted in compliance with the Helsinki Declaration, and ethical approval was granted by the Cambridgeshire 2 Research Ethics Committee. For the current study ethical approval was granted by the University of Glasgow, College of Medical, Veterinary & Life Sciences Ethics Committee (Appendix 2.4).

Results

Sample Characteristics

Table 2.1 includes all baseline characteristics of the normative sample. The final sample consisted of 648 participants of which 49.8% were male. The mean age was 54.5 years (sd=18.3) and age ranged from 18 to 88 years. The sample consisted predominantly of white British people and education level was high with 61% of participants having a university degree or higher and an average of 15 (sd=3.3) years in education with a range of 9 to 22 years in total.

Table 2.1
Demographic Characteristics

N=648	
Age (Mean, Sd)	54.5, 18.3
Education (Mean Sd)	15, 3.3
Gender	
Male	49.8%
Female	50.2%
Nationality	
British/	93%
Other	7%
Ethnic Group	
White	96%
Other	4%
Highest Qualification	
University or Higher	61%
A/O Levels	22.5%
None	6.5%

Table 2.3 summarises the scores on all neuropsychological tests. It is worth noting that the mean IQ, as measured by the STW scaled score, (m=12.62, sd= 2.9) corresponds to a standard score of approximately 110, which would be slightly higher than the population average. Moreover, seven participants scored below the cut-off of 80 (range 76-79) on the ACE-R. However, as participants had already been screened for neurodegenerative conditions using different screening tools (see Shafto et al. 2014) and because participants with low scores tended to be older, this was considered potentially a sign of natural cognitive decline due to

aging that could be encountered in the general population. It was therefore, decided to include the participants in the analysis, to achieve a sample more representative of the general population.

Table 2.2
Descriptive Statistics for neuropsychological measures

Measure	Mean	SD	Range	
			Minimum	Maximum
Hotel Time Deviation*	270.23	250.57	20.19	813.09
Hotel No of Tasks*	5	1	2	5
STW scaled	12.62	2.97	1	18
WMS-III Immediate	14.64	3.96	3	24
WMS-III Delayed	12.96	4.19	0	24
Cattell	31.78	6.78	11	44
Proverbs*	5	2	0	6
ACE-R*	96	5	76	100
STR*	.34	.08	.24	1.69
CTR*	.54	.16	.34	1.4
HADS-Anxiety*	4	4	0	15
HADS-Depression*	2	3	0	15

Abbreviations: No= number; STW= Spot the Word; WMS= Wechsler Memory Scale; ACE- R= Addenbrooke's Cognitive Examination-Revised; STR= Response Time Simple; CTR= Response Time Choice; HADS= Hospital Anxiety and Depression Scale.
Note: *Median and interquartile range are presented for the non-normally distributed variables.

Regression Analysis

Multiple regression was conducted to examine the effect of the predictor variables on the Square Root transformed Hotel Test Time Deviation score. Sex did not have a significant effect and was subsequently removed from further analyses. AgeC, age^2, years in education and IQ were all found to have a significant effect on performance on the Hotel Test Time deviation score (F (4,640) = 18.992, p < .001, R^2 = .106). Overall, the model explained 10% of the variance. Hotel scores tended to increase with age and decrease with years of education and higher IQ scores. The final regression model can be found in Table 2.3.

Table 2.3

Multiple regression model for Square Root transformed Hotel Test Time Deviation

	B	**SE B**	**t**	**p**	$\mathbf{R^2}$	**SDe (Resid)**
Constant	21.131	1.041	20.295	<.001	.106	4.634
AgeC	.063	.011	5.994	<.001		
Age^2	.001	.001	2.252	.025		
STW (IQ)	-.224	. 068	-3.320	<.001		
Years/Education	-.135	.064	-2.128	.034		

Notes: SDe (Resid)= Residuals' standard error; Agec = (Age-54); Age^2 = ($AgeC^2$); STW= Spot the Word; IQ= intelligence quotient.

Establishing regression-based norms

The regression equation including all significant predictors needs to be used to obtain a percentile range and a standardised score for any given Hotel Test Time Deviation score. An example will be explained below to illustrate the process of obtaining the norms for individual scores. Clinicians and researchers can use the supplementary calculator located in Appendix 2.1.

The example is of a 48-year-old male with 17 years in education and an IQ scaled score of 14, who scores 410.5 seconds on the Hotel Test time deviation score. First, as data was transformed the square root of the score needs to be established which is 20.26. The mean-centred age for this individual would be 48-54= -6 years. Knowing the important predictors, this person's predicted score could be calculated using the equation: $Y = B_0 + (B_1 \times AgeC) + (B_2 \times Age^2) + (B_3 \times IQ) + (B_4 \times Education)$. From the information contained in Table 2.3 the equation then translates to: $Y = 21.131 + (.063 \times -6) + (.001 \times 36) + (-.224 \times 14) + (-.135 \times 17) = 15.36$. The residual value can then be calculated as: Re = Observed Score (Square Root of Time Deviation) – Predicted Score = 20.26 – 15.36= 4.9. After consulting Table 2.3 again, the residual value can be standardised using the SDe value: Z= Re/SDe= 4.9/4.634= 1.06. Because higher hotel scores indicate worse performance we insert (-) on the SDe value to acquire the correct percentile score. Therefore SDe= -1.06, indicating that this participant scores 1.06 deviations below the average. Finally, using the normal cumulative distribution it can be established that this score is in the 15^{th} percentile which corresponds to the Low Average range.

Similarly, it is illustrated below how clinicians can use the RegBuild_MR_Raw.exe computer programme to produce the norms for the same case (Crawford et al., 2012). Once clinicians have accessed the raw data (Appendix 2.2), the individual scores are added in the last row of the file. Appendix 2.3 explains the process of inputting data into the programme. Using the same example of the 48-year-old male the output of the programme which is in Appendix 2.5 will be explained.

The output first provides a summary of the data followed by means and standard deviations as well as the regression model for the data provided. In addition, the next section includes the results of the regression-based norming procedure described above. As in the example above the case's observed score (Square root of the time deviation score) is 20.26 whilst the predicted score is 15.46. The discrepancy between the two scores is 4.7, indicating that according to the regression model this individual scores 4.7 higher than what it would be predicted by the model. The effect size (Z-OP) for the difference between the observed and the predicted score corresponds to the standardised residual value and according to the output is Z-OP= 1.03 (95% CI = .87 – 1.17) suggesting that the 48-year-old male performed approximately 1 Sd above the normative prediction. Note that for the Hotel Test time deviation score, higher values indicate greater deficit.

Next the programme evaluates the t score value, in this case $t(494)= 1.02$, $p= .15$ and therefore not significantly below the score predicted from his baseline score.

Finally, the output provides an estimate of the percentage of the population who would demonstrate a discrepancy more extreme than the one analysed = 15.3% as well as the '95% confidence limits on the percentage' which equals 11.95% to 19.10%. In other words, this is an estimate of how rare this discrepancy is. This suggests that around 15% of the normal sample had a difference of this size between observed and estimated score. Thus, although the discrepancy is not statistically significant, the size of this discrepancy is estimated to be encountered in 15% or less of the adult normal population which would place this individual in the Low Average range.

Hotel Number of Tasks

As mentioned above the Hotel Number of Tasks score was highly skewed due to the presence of a ceiling effect. Therefore, regression-based norms were not calculated. Table 2.4 includes the percentile range for the Hotel Number of tasks scores stratified by age.

Table 2.4

Percentiles by age for Number of Tasks

		Percentile Range			
Age Range					
	Hotel Number of Tasks	**2**	**3**	**4**	**5**
18-50		<5th	5th -15th	15th – 40th	>40th
51-65		<5th	5th – 20th	20th – 45th	>45th
66-80		<10th	10th – 30th	30th – 65th	>65th
>80		<10th	20th -40th	40th – 70th	>70th

Note: As completion of only one task is indicative of failing to understand the instructions, calculating a percentile equivalent is not recommended.

Validity

Spearman's rank correlations (rho) were calculated to explore the relationship between performance on the Hotel Test and different neuropsychological tests. Table 2.5 provides a summary of correlations for all variables.

Significant but small negative correlations were found between the Hotel Test time deviation score and performance on the Cattell test and Proverbs test. Similar results were observed for the relationship between the Hotel time deviation score and performance on the ACE-R and the WMS logical memory subtest, as well as performance on the reaction time tasks.

Table 2.5

Spearman Correlations Matrix

	LM/Imm[a]	LM/Del[a]	Cattell	Proverbs	ACE-R	STR	CTR
Hotel Time deviation	-.108**	-.132**	-.255**	-.109**	-.227**	.151**	.183**
	N=647	N=647	N=647	N=625	N=647	N=603	N=600
Hotel No of Tasks	.131**	.169**	.259**	.088*	.232**	-.122**	-.165**
	N=647	N=647	N=647	N=625	N=647	N=603	N=600

Abbreviations: LM/Imm= Logical Memory Immediate; LM/Del= Logical Memory Delayed; ACE- R= Addenbrooke's Cognitive Examination-Revised; STR= Response Time Simple; CTR= Response Time Choice; No= number.

Note: [a]Pearson's Correlations are reported for correlations between Cattell scores and WMS-III scores, $**p<.001$; $*p<0.05$

In addition, small but significant positive correlations were also observed between the Number of tasks attempted on the Hotel Test and all other variables. All correlations for the Hotel Test sub scores were significant at $p< .001$, however given the large sample, more attention should be paid on the magnitude of the effect.

Discussion

The study aimed to provide continuous norms from the UK population for the Hotel Test, a test of multitasking (Manly et al., 2002). The norms are based on a sample of 648 healthy adults and can be used to enhance clinical practice in the assessment of EF. Neuropsychological test norms have particular significance for clinical practice as they allow clinicians to directly contrast individual patients' performance with the general population whilst considering relevant factors such as age. It should be noted that norms are presented for the modified version of the Hotel Test. Even though a prospective memory element (garage doors) has been removed it is thought that this element is built in other parts of the test, as participants need to remember to switch between tasks. Moreover, it could be argued that the garage door component could act as a cue for people to check the clock, prompting them to move on to a different task and thus affecting performance. Additionally, the modified version is shorter and incorporates more easily accessible materials, thus could be easier to use in clinical settings.

The results of the current study indicate a significant effect of age, education and IQ as assessed by the STW test on multitasking performance. Biological sex on the other hand, did not predict performance on the test.

With respect to age the results indicate a positive linear effect of age of EF performance as well as a quadratic effect, albeit small. This is consistent with studies of developmental trajectories of EF which indicate that EF function increases into early adulthood and subsequently declines in later middle to old age (Clark et al., 2006). The small quadratic effect observed in this study probably reflects the lack of children in the sample and the presence of a proportion of young adults. Clark et al. (2006), observed a similar quadratic function of age where the EF performance asymptote was reached in the 20–29-year-old age group. Overall, there seems to be consensus within the literature for marked changes in EF compared to other abilities as age progresses. It has been suggested that age-related changes are due to frontal degeneration and this is supported by neuroimaging studies which indicate an earlier deterioration of the frontal area compared to other brain areas (Allain et al., 2005; Raz et al., 2005). Indeed, Andrés & Van der Linden (2000) demonstrated that older participants were considerably more impaired on tasks of planning, inhibition, and abstraction of logical rules compared to younger people. Similarly, and with reference to multitasking ability, it has been shown that performance on tests of multitasking such as the MET and crucially the SET (on which the Hotel Test was based) also decreases with age (Allain et al., 2005; McAllister & Schmitter-Edgecombe, 2013). The results of this study confirm those of past research.

The study also found a significant effect of education, as measured by years in education with the Hotel Test Time deviation scores decreasing for people with higher educational level. McAllister & Schmitter-Edgecombe (2013) also reported that level of education as assessed by years in education emerged as a unique predictor of multitasking ability on a naturalistic assessment of EF. In addition, Diaz-Asper et al. (2004) divided 220 healthy adults in three educational level categories according to years in education. They demonstrated that individuals in the below-average group performed significantly worse in tests of EF than people in the average group and individuals in the average group performed worse than those in the above-average. In the same sample, they also noted that IQ scores as measured by the Wechsler Adult Intelligence Scale also predicted performance on neuropsychological tests of EF. This is consistent with findings of this study which reported a significant effect of IQ score as measured by the STW on Hotel Test time deviation scores. In fact, IQ and educational level were stronger predictors of Hotel Test performance than age. Similarly, EF has also been found to predict academic attainment and crystallised IQ suggesting a strong link between the two (Brydges at al., 2012).

Our findings regarding the effect of sex are in accordance with previous research which also failed to find such an effect. For example, McAllister & Schmitter-Edgecombe (2013) assessed EF in healthy younger and older adults using a naturalistic task of multitasking and found no significant effects of sex. More recently, in their meta-analysis Gaillard et al. (2021) reported that males and females did not differ in their performance on three EF control domains, namely performance monitoring, response inhibition and cognitive set shifting. Although some sex differences have been reported regarding EF performance, there is growing consensus that biological sex does not impact of performance of EF tests. (Clark et al., 2006). Indeed, Grissom & Reyes (2019) evaluated the work on EF sex differences in humans and animals and concluded that there is little support for the idea that gender significantly impacts on EF.

The use of regression-based norms is becoming an increasingly popular alternative to traditional discrete norming procedures and has been applied to a variety of neuropsychological tests (Walker et al., 2017). Whilst easier to use and calculate, discrete norms need large samples to calculate norms precisely (Oosterhuis et al., 2016). For example, in their study Oosterhuis et al. (2016) demonstrated that to achieve the same level of precision, a sample size of between 100 and 500 participants would be needed for continuous norming compared to approximately 4000 for discrete norming procedures. Additionally, by using regression models to evaluate the role of different predictors which are treated as continuous, this approach utilises the whole sample instead of arbitrary cut-off points in the variables of interest (Lenhard & Lenhard, 2020). This is particularly important as it has been suggested that arbitrarily dividing continuous variables such as age and IQ into normative groups may significantly affect score interpretation when moving between groups, depending on which group individuals might fall into (Oosterhuis et al., 2016). Finally, this approach is particularly useful as it allows for the estimation of linear and non-linear effects of relevant factors thus allowing for more precision in score prediction (Van Breukelen & Vlaeyen, 2005).

In terms of the Hotel Test Number of tasks score, data was highly skewed due to ceiling effects and therefore continuous norms could not be calculated. Percentiles stratified by age are presented for reference, confirming that although suboptimal performances on this score might be a good indicator of EF deficit overall, the score is not very sensitive to smaller nuances in deficits. The observed ceiling effects for this score could prevent clinicians from effectively using this score due to the limited sensitivity as it might allow patients with mild EF impairment to perform at maximum or near maximum level.

Moreover, the secondary aim of the study was to provide information regarding the validity of the Hotel Test by exploring correlations with performance on associated neuropsychological tests. Overall, we found significant but small associations between both the Hotel Test time deviation and number of tasks scores and measures of fluid intelligence, verbal memory, general cognitive functioning as well as a measure of EF, psychomotor speed and speed of processing and inhibition. The highest correlation observed, albeit still small, for both Hotel Test scores was observed with the Cattell Culture Fair Test (Cattell, 1971). A similar correlation between the Hotel Test and the Cattell test was also reported by Roca et al. (2010). This is in accordance with recent conceptualisations of EF which suggest that executive functions although related, are not the same as general intelligence (g; Friedman & Miyake, 2017).

In addition, small but significant correlations were observed between the Hotel Test sub scores and measures of verbal memory and general cognitive function demonstrating good divergent validity. Moreover, the small correlations between Hotel Test and RT scores indicate that performance on this test of multitasking is not heavily influenced by slow processing and motor speed. Although it could be argued that small correlations might indicate lack of convergent validity the tests explored in this study represented conceptually relevant but different constructs to that of EF (e.g verbal memory) and therefore a higher correlation was not anticipated. Interestingly, the smallest correlations were observed between the Hotel Test scores and the Proverbs score, another test of EF also linked to anterior frontal function and associated with set shifting, inhibition and abstract thinking (Shafto et al., 2014). Although perhaps surprising at first, it is worth emphasising that correlations between EF tests have frequently been medium to small and often not statistically significant (e.g., Ardila, 2018; Torralva et al., 2012). Such findings support the idea that EF does not represent a unitary factor but rather a multitude of related but diverse underlying processes (Ardila, 2018; Friedman & Miyake, 2017). Equally, Jovanovski et al. (2012) found no significant corelations between EF tests and their naturalistic test of multitasking. The only significant correlation was reported for the modified version of the SET, which also did not correlate with any other measures of EF. A possible explanation for the finding is that more unitary measures of EF such as Proverb comprehension measure distinct EF processes. On the contrary, multitasking tests require the ability to employ these EF processes whilst monitoring performance and updating and adjusting behaviour (Shallice & Burgess, 1991).

Strengths and Limitations

The current study has several strengths. It presents normative data and preliminary validity information for an easily administered and ecologically valid test of multitasking. The study includes a large sample of healthy participants, also making it a good study for some preliminary exploration of the association of the Hotel Test with other neuropsychological tests. Additionally, the rigorous sampling procedures allowed for good representation and spread of demographic variables in order to address potential bias associated with the effect of these factors on the Hotel Test performance. Additional strengths include the use of regression-based norming procedures which allow for adjustment of relevant confounding effects. The inclusion of an accessible calculator and the use of a freely accessible programme to assist clinicians in generating norms, rather than asking them to do this by manual process also increases clinical utility and minimises the potential for human error.

However, the study should be assessed in light of its limitations. The normative sample had a higher-than-average IQ score as assessed by the STW. Moreover, based on census data, more people in the sample identified as White/British compared to what is reported on a national level (93% compared to 78%; Office of National Statistics, 2021). Although regression-based norms can be applied for a wide range of intellectual abilities, some caution is needed when applying the norms. Moreover, IQ was measured using the STW scaled scores. Although by using a psychometric conversion table different IQ scores can be converted into scaled scores and utilised in clinical practice, inclusion of a more commonly used test of intelligence might have enhanced clinical utility. As this was a secondary data analysis study however, this was not possible.

Conclusions and Future Directions

The study builds on previous research and provides evidence on the use of a measure of multitasking namely the Hotel Test. The use of regression-based norms increases the test's clinical utility by allowing clinicians to estimate performance accounting for age, IQ and education. However, given the sparsity of research utilising the tool further research is needed to assess the psychometric properties and clinical effectiveness of the Hotel Test. More specifically, more evidence is needed as it pertains to the psychometric properties of the test and especially the validity of the Hotel Test. Finally, as norms are reported for the modified version of the test is important to understand how performance on the two versions compare.

References

American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing. (1999). *Standards for educational and psychological testing*, USA.

Andrés, P., & Van der Linden, M. (2000). Age-related differences in supervisory attentional system functions. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *55*(6), P373-P380.

Ardila, A. (2018). Is intelligence equivalent to executive functions?. *Psicothema*, 30(2), 159-164.

Army Individual Test Battery. (1944). Manual of directions and scoring. Washington, DC: War Department, Adjutant General's Office.

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology*, *32*(1), 55-65.

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and operationalization of executive function. *Mind, Brain, and Education*, *10*(1), 10-33.

Banich, M. T. (2009). Executive function: The search for an integrated account. *Current directions in psychological science*, *18*(2), 89-94.

Brydges, C. R., Reid, C. L., Fox, A. M., & Anderson, M. (2012). A unitary executive function predicts intelligence in children. *Intelligence*, 40, 458-469.

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L. A., Dawson, D. R., Anderson, N. D., Gilbert, S. J., Dumontheil, I. & Channon, S. (2006). The case for the development and use of" ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the international neuropsychological society*, *12*(2), 194-209.

Cattell, R. B. Abilities: their Structure, Growth, and Action (Houghton-Mifflin, 1971).

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of clinical neuropsychology*, *23*(2), 201-216.

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of clinical neuropsychology*, *21*(3), 217-227.

Clark, C. R., Paul, R. H., Williams, L. M., Arns, M., Fallahpour, K., Handmer, C., & Gordon, E. (2006). Standardized assessment of cognitive functioning during development and aging using an automated touchscreen battery. *Archives of Clinical Neuropsychology*, *21*(5), 449-467

Crawford, J. R., Garthwaite, P.H., Denham, K.A., & Chelune, G.J. (2012). Using regression equations built from summary data in the psychological assessment of the individual case: extension to multiple regression. *Psychological Assessment, 24*, 801-814.

Diaz-Asper, C. M., Schretlen, D. J., & Pearlson, G. D. (2004). How well does IQ predict neuropsychological test performance in normal adults?. *Journal of the International Neuropsychological Society*, *10*(1), 82-90.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186-204.

Gaillard, A., Fehring, D. J., & Rossell, S. L. (2021). A systematic review and meta-analysis of behavioural sex differences in executive control. *European Journal of Neuroscience*, *53*(2), 519-542

Gioia, G. A., & Isquith, P. K. (2004). Ecological assessment of executive function in traumatic brain injury. *Developmental neuropsychology*, *25*(1-2), 135-158.

Goldstein, S., & Naglieri, J. A. (2014). Handbook of Executive Functioning. New York, NY: Springer.

Gracey, F., Fish, J. E., Greenfield, E., Bateman, A., Malley, D., Hardy, G., Ingham, I., Evans, J. J. & Manly, T. (2017). A randomized controlled trial of assisted intention monitoring for the rehabilitation of executive impairments following acquired brain injury. *Neurorehabilitation and neural repair*, *31*(4), 323-333.

Grant, D.A., & Berg, E.A. (1948). A behavioural analysis of degree or reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology, 38*, 404–411.

Grissom, N. M., & Reyes, T. M. (2019). Let's call the whole thing off: evaluating gender and sex differences in executive function. *Neuropsychopharmacology*, *44*(1), 86-96.

Hodges J.R (1994). Neurological aspects of dementia and normal aging. In *Dementia and Normal Aging.*Cambridge University Press, 118–129.

IBM Corp. (2021). IBM SPSS statistics for Windows. 28.0 ed, New York: IBM Corp.

Jansari, A. S., Devlin, A., Agnew, R., Akesson, K., Murphy, L., & Leadbetter, T. (2014). Ecological assessment of executive functions: a new virtual reality paradigm. *Brain Impairment*, *15*(2), 71-87.

Jovanovski, D., Zakzanis, K., Campbell, Z., Erb, S., & Nussbaum, D. (2012). Development of a novel, ecologically oriented virtual reality measure of executive function: The Multitasking in the City Test. *Applied Neuropsychology: Adult*, *19*(3), 171-182.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). Applied linear statistical models (5th ed.).NewYork: McGraw Hill.

Lamberts, K., Evans, J., & Spikman, J. (2010). A real-life, ecologically valid test of executive functioning: The executive secretarial task. *Journal of Clinical & Experimental Neuropsychology*, *32*(1), 56–65.

Lenhard, W., & Lenhard, A. (2020). Improvement of Norm Score Quality via Regression-Based Continuous Norming. *Educational and Psychological Measurement*, 81(2), 229–261.

Manly, T., Hawkins, K., Evans, J., Woldt, K., & Robertson, I. H. (2002). Rehabilitation of executive function: facilitation of effective goal management on complex tasks using periodic auditory alerts. *Neuropsychologia*, *40*(3), 271-281.

McAlister, C., & Schmitter-Edgecombe, M. (2013). Naturalistic assessment of executive function and everyday multitasking in healthy older adults. *Aging, Neuropsychology, and Cognition*, *20*(6), 735-756.

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, *21*(11), 1078-1085.

Mole, J. A., & Demeyere, N. (2020). The relationship between early post-stroke cognition and longer-term activities and participation: A systematic review. *Neuropsychological rehabilitation*, *30*(2), 346-370.

Nalder, E. J., Clark, A. J., Anderson, N. D., & Dawson, D. R. (2017). Clinicians' perceptions of the clinical utility of the Multiple Errands Test for adults with neurological conditions. *Neuropsychological rehabilitation*, *27*(5), 685-706.

Office of National Statistics. (2021). *Producing admin-based ethnicity statistics for England: methods, data and quality.* Retrieved from: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/producingadminbasedethnicitystatisticsforenglandmethodsdataandquality/2021-08-06.

Oosterhuis van der Ark, H. E., L. A., & Sijtsma, K. (2016). Sample size requirements for traditional and regression-based norms. *Assessment*, *23*(2), 191-202.

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., Dahle, C., Gerstorf, D., & Acker, J. D. (2005). Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. Cerebral Cortex, 15(11), 1676-1689.

Rivera, D., & Arango-Lasprilla, J. C. (2017). Methodology for the development of normative data for Spanish-speaking pediatric populations. *NeuroRehabilitation*, *41*(3), 581-592.

Roca, M., Manes, F., Chade, A., Gleichgerrcht, E., Gershanik, O., Arévalo, G. G., Torralva, T. & Duncan, J. (2012). The relationship between executive functions and fluid intelligence in Parkinson's disease. *Psychological medicine*, *42*(11), 2445-2452.

Roca, M., Manes, F., Gleichgerrcht, E., Watson, P., Ibáñez, A., Thompson, R., Torralva, T. & Duncan, J. (2013). Intelligence and executive functions in frontotemporal dementia. *Neuropsychologia*, *51*(4), 725-730.

Roca, M., Parr, A., Thompson, R., Woolgar, A., Torralva, T., Antoun, N., Manes, F. & Duncan, J. (2010). Executive function and fluid intelligence after frontal lobe lesions. *Brain*, *133*(1), 234-247.

Shafto, M. A., Tyler, L. K., Dixon, M., Taylor, J. R., Rowe, J. B., Cusack, R., Calder, A. J., Marslen-Wilson, W. D., Duncan, J., Dalgeish, T., Henson, R. N., Brayne, C., Cam-CAN & Matthews, F. E. (2014). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC neurology*, *14*(1), 1-25.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B, 298*, 199–209.

Shallice, T., and Burgess, P. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain 114*, 727–741.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Taylor, J. R., Williams, N., Cusack, R., Auer, T., Shafto, M. A., Dixon, M., Tyker, L., Cam-CAN & Henson, R. N. (2017). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *Neuroimage*, *144*, 262-269.

Torralva, T., Gleichgerrcht, E., Lischinsky, A., Roca, M., & Manes, F. (2013). "Ecological" and highly demanding executive tasks detect real-life deficits in high-functioning adult ADHD patients. *Journal of Attention Disorders*, *17*(1), 11–19.

Torralva, T., Roca, M., Gleichgerrcht, E., Bekinschtein, T., & Manes, F. (2009). A neuropsychological battery to detect specific executive and social cognitive impairments in early frontotemporal dementia. *Brain*, *132*(5), 1299-1309.

Torralva, T., Strejilevich, S., Gleichgerrcht, E., Roca, M., Martino, D., Cetkovich, M., & Manes, F. (2012). Deficits in tasks of executive functioning that mimic real-life scenarios in bipolar disorder. *Bipolar Disorders*, *14*(1), 118-125.

Van Breukelen, G. J., & Vlaeyen, J. W. (2005). Norming clinical questionnaires with multiple regression: the Pain Cognition List. *Psychological assessment*, *17*(3), 336.

Walker, L. A., Marino, D., Berard, J. A., Feinstein, A., Morrow, S. A., & Cousineau, D. (2017). Canadian normative data for minimal assessment of cognitive function in multiple sclerosis. *Canadian Journal of Neurological Sciences*, *44*(5), 547-555.

Weschler CJ. (1999) *Wechsler Memory Scale. Third UK edition.* London: Harcourt Assessment.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361-370.

Appendices

Appendix 1.1 Search Strategies for Systematic review by database

CINAHL:

Print Search History: EBSCOhost 21/09/2021, 13:02

MY EBSCOhost

Tuesday, September 21, 2021 12:02:16 PM

#	Query	Limiters/Expanders	Last Run Via	Results
S4	S1 AND S2	Limiters - Published Date: 20010101-20211231 Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - CINAHL	19
S3	S1 AND S2	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - CINAHL	20
S2	TX "executive function*" OR TX multitask*	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - CINAHL	16,751
S1	TX Hotel*	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - CINAHL	8,829

https://web.b.ebscohost.com/ehost/searchhistory/PrintSearchHi...VN0YW5kYXJkJnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d&theSearchHistoryIds= Page 1 of 1

PsychInfo:

Tuesday, September 21, 2021 10:48:14 AM

#	Query	Limiters/Expanders	Last Run Via	Results
S3	S1 AND S2	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - APA PsycInfo	40
S2	TX "*executive function*" OR TX multitask*	Limiters - Publication Year: 2001-2021 Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - APA PsycInfo	36,247
S1	TX Hotel*	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - APA PsycInfo	7,875

Medline through EBSCOhost:

Tuesday, September 21, 2021 10:53:21 AM

#	Query	Limiters/Expanders	Last Run Via	Results
S4	S1 AND S2	Limiters - Date of Publication: 20010101-20211231 Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE	37
S3	S1 AND S2	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE	39
S2	TX "*executive function*" OR TX multitask*	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE	42,872
S1	TX Hotel*	Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE	20,502

Medline search through OVID:

Ovid MEDLINE(R) <1996 to September Week 2 2021>

1 Hotel*.mp. 2728
2 ("executive *function*" or multitask*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] 32538
3 1 and 2 10

Embase:

Embase <1996 to 2021 Week 37>

1 Hotel*.mp. 4563
2 ("executive *function*" or multitask*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword heading word, floating subheading word, candidate term word] 66386
3 1 and 2 31

Pro Quest:

Name: 21/09/2021Edit name
Searched for: (((Hotel NEAR/1 (test OR task)) AND ((executive function*) OR (multitask*))) AND stype.exact("Scholarly Journals" OR "Dissertations & Theses") AND la.exact("English") AND pd(>20011231)) NOT subt.exact("management" OR "hotels & motels" OR "marketing" OR "business administration" OR "recreation" OR "american history" OR "employees" OR "hospitality industry" OR "labor relations" OR "accounting" OR "customer services" OR "linguistics" OR "tourism" OR "womens studies" OR "american studies" OR "human resource management" OR "job satisfaction" OR "profitability" OR "sociology" OR "art history" OR "brand equity" OR "brand loyalty" OR "competition" OR "consumers" OR "consumption" OR "customer satisfaction" OR "economic growth" OR "economic impact" OR "economics" OR "finance" OR "food science" OR "hotels" OR "international relations" OR "market strategy" OR "information technology" OR "organizational behavior" OR "social structure" OR "african history" OR "annual meetings" OR "annual reports as topic" OR "batteries" OR "children" OR "children & youth" OR "civil engineering" OR "councils" OR "environmental economics" OR "environmental management" OR "financial reporting") AND pd(>20011231)
Limited by:
Date: After 2001
Databases: 20 databases searched

Scopus:

(TITLE-ABS-KEY ((hotel W/1 (test OR task))) AND TITLE-ABS-KEY (("*executive" OR multitask*))) 17 results

Web of science:

3 (#1) AND #2.
15

2 TS=("executive function*" OR multitask*) OR AB=("executive function*" OR multitask*)
75,009

1 TS=(hotel NEAR/1 (test OR task)) OR AB=(hotel NEAR/1 (test OR task)). 155

Appendix 1.2. Summary of Methodological Quality

Methodological Quality Assessment of included studies				
Property	Study	Overall rating	Area with Lowest Rating	Main Limitations
Validity: Convergent	Baylan, 2014	Doubtful	Design	Limited information on the measurement properties of the comparator instrument on any population
	Baylan, 2014	Very good	n/a	
	Cullen et al., 2016	Adequate	Design	Limited information on the measurement properties of the comparator instrument on ABI population
	Cullen et al., 2016	Very good	n/a	
	Fish, 2008	Doubtful	Stats	Limited information on measures of dispersion. Unclear if data are normally distributed to justify use of parametric test. No correction for multiple comparisons. Combined scores of second administration where practice effects might have been evident.
	Fish et al., 2007	Very good	n/a	
	Roca et al., 2010	Very Good	n/a	
Validity: Ecological	Baylan, 2014	Very Good	n/a	
	Baylan, 2014	Inadequate	Design	No information on the measurement properties of the comparator instrument on any population.
	Fish et al., 2007	Doubtful	Design	Limited information on the measurement properties of the comparator instrument on any population. Unclear why only results for Hotel number of tasks is reported.

Validity Known group	Banks et al., 2016	Doubtful	Design/ Stats	Although there is some description of ten groups important information is missing with regard to education, premorbid IQ & other cognitive characteristics. No information on whether groups where matched on those. Large standard deviations reported for the patient group but no mention if data was checked for normality.
	Baylan, 2014	Doubtful	Design	Limited information on differences between groups especially for cognitive characteristics. Significant group differences in premorbid IQ, gender and MH symptoms.
	Fish, 2008	Doubtful	Design	Limited information on sample characteristics, matched only on age reported not sex or IQ. Unclear information on data distribution.
	Manly et al., 2002	Very good	n/a	
	Roca et al., 2010	Adequate	Design/ Stats	Information missing on participants' gender. Unclear if groups were matched for age, sex and premorbid IQ. Unclear information on data distribution
	Roca et al., 2011	Adequate	Stats	Unclear information on data distribution
Responsiveness	Gracey et al., 2017	Inadequate	Design/ Stats	Unclear which Hotel Test score is used and reported for the analysis. Significant differences between groups. High drop-out rates and study not powered for subsidiary analysis.
	Levine et al., 2011	Doubtful	Design/ Stats	No information on power calculation. Sample not exclusively randomly allocated
	Tornas et al., 2016	Very Good	n/a	n/a

Note. Studies reported twice have been rated separately assessed and the result was different

Appendix 1.3. Rating per psychometric property

Convergent Validity								
Study	**Lowest Quality Score/ Category**	**Sample (n, group)**	**Comparator Measure**	**Results: Correlation between Hotel scores and comparator measures (r=)**				**Rating[a] (+/-/?)**
				No of Tasks	**Time Dev**	**GD**	**GD/Dev**	
Baylan, 2014	Doubtful/ Design	39, ABI	Computerised tests	-.21 ns (Spearman)	.22 ns	n/a	n/a	(-)
Baylan, 2014	Very Good		CAMPROMPT	.49 ** (Spearman)	-.51**	n/a	n/a	(+)
Cullen et al., 2016	Adequate/ Design	18, ABI	CMET	0.67**	n/a	n/a	n/a	(+)
Cullen et al., 2016	Very Good		Modified SET	0.39*	n/a	n/a	n/a	(+)
Fish, 2008	Doubtful/ Stats	16, ABI	RT	n/a	-.23			(-)
			VF	n/a	-.67*			(+)
			DS-B	n/a	-.14			(-)
			VESPAR	n/a	-.70*			(+)
			Hay-A.	n/a	.13			(-)
			Hay-B	n/a	.19			(-)
			SART (Com error)	n/a	.12			(-)
			PM	n/a	-.56*			(+)
Fish et al., 2007	Adequate/ Stats	20, ABI	SART error rate	−.61** (Spearman)	n/a	n/a	n/a	(+)
Roca et al., 2010	Very Good	20, ABI 25, HC	Culture fair	n/a	.25*	n/a	n/a	(+)

Pooled Data	No of tasks:77 Time dev: 100	+ Inconsistent ±	No tasks: 4/5 (80%) confirmed Time Dev: 5/11 (45%) confirmed	Sufficient

Abbreviations: No= number; Time Dev= Time deviation; GD= Garage Doors; ABI= Acquired Brain Injury; ns= non-significant; n/a= not applicable; CAMPROMPT= Cambridge Test of Prospective Memory; CMET= Computerised Multiple Elements Test; SET= Six Elements Test; RT= Response Time; VF= Verbal Fluency; DS-B= Digit-span backwards; VESPAR= Verbal Composite Scale; Hay-A= Hayling Part A time; Hay-B= Hayling Part B time; SART= Sustained Attention to Response Test; Com= Commission; PM= Prospective Memory Task; HC= Healthy Control.

Note: [a](+) sufficient, (?) indeterminate, (-) insufficient, or (?) mixed; * $p \leq .05$; ** $p \leq .01$;

Ecological Validity								
Study	**Lowest Quality Score/ Category**	**Sample (n, group)**	**Comparator Measure**	**Results: Correlation between Hotel scores and comparator measures (r=)**				**Rating[a] (+/-/?)**
				No of Tasks	**Time Dev**	**GD**	**GD/Dev**	
Baylan, 2014	Very Good	39, ABI	PRMQ	.35* (spearman)	.37* (spearman)	n/a	n/a	(+)
Baylan, 2014	Inadequate/ Design		GMQ	.29 ns (spearman)	.29 ns (spearman)	n/a	n/a	(-)
Fish et al., 2007	Doubtful/ Design	20, ABI	PM task	−.043 ns	n/a	n/a	n/a	(-)
Pooled Data		No of tasks: 59 Time dev: 39		± Sufficient +	No tasks: 1/3 (66.6%) confirmed Time Dev: 1/2 (50%) confirmed			Inconsistent

Abbreviations: No= number; Time Dev= Time deviation; GD= Garage Doors; ABI= Acquired Brain Injury; PRMQ= Prospective and Retrospective Memory Questionnaire; ns= non-significant; n/a= not applicable; GMQ= Goal Management Questionnaire; PM= Prospective Memory

Note: [a](+) sufficient, (?) indeterminate, (-) insufficient, or (?) mixed; * $p \leq .05$

Known-Group Validity								
Study	**Lowest Quality Score/ Category**	**Sample (n, group)**		**Between group differences on Hotel Scores (direction and Cohen's d)**				**Criteria[a] (+/-/?)**
		Group 1	**Group 2**	**No of Tasks**	**Time Dev**	**GD**	**GD/DV**	
Banks et al., 2016	Doubtful/ Design, Stats	13, mTBI	11, HC	n/a	HC Lower 0.8 ns	n/a	n/a	(+)
Baylan, 2014	Doubtful/ Design	39, ABI	16, HC	ABI Lower, 0.6*	HC Lower, 0.68*	n/a	n/a	(-)
Fish, 2008	Doubtful/ Design	16, ABI	12, HC	ABI Lower, 1.32*	HC Lower, 1.25*	n/a	n/a	(+)
Manly et al., 2002	Very Good	10, TBI	24, HC	ABI Lower, 0.18**	HC Lower, 1.88**	n/a	HC Lower, 0.9	(+)
Roca et al., 2010	Adequate/ Design, Stats	20, ABI tumour	25, HC	n/a	HC Lower, 1.14**	n/a	n/a	(+)
Roca et al 2011	Adequate/ Stats	7, BA10 lesions	25, HC	BA Lower, 1.44**	HC Lower, 1.24*	n/a	n/a	(+)
Pooled Data		No of tasks: 149 Time dev: 218				No of tasks: 2/4 (50%) ± Time Dev: 5/6 (83%) +		Inconsistent Sufficient

Abbreviations: No= number; Time Dev= Time deviation; GD= Garage Doors; mTBI= mild Traumatic Brain Injury; HC= Healthy Control; ns= non-significant; n/a= not applicable; ABI= Acquired Brain Injury; TBI= Traumatic Brain Injury; BA= Brodmann area .

Note: [a](+) sufficient, (?) indeterminate, (-) insufficient, or (?) mixed; * p ≤ .05; ** p ≤ .01;

Responsiveness										
Study	**Lowest Quality Score/ Category**	**Design**	**Intervention**	**Groups (n, group)**		**Within-group differences on Hotel Scores (Cohen's d)**				**Rating[a] (+/-/?)**
				Intervention	**Control**	**No of Tasks**	**Time Deviation**	**GD**	**GD/DV**	
Gracey et al., 2017	Inadequate/ Other	RCT with cross-over	Assisted Intention Monitoring vs Brain education and games	32, ABI	30, ABI	d= 0.06 ns between-group[b]		n/a	n/a	(?)
Levine et al., 2011	Doubtful/ Design, Other	Partial RCT	GMT/ Brain Health workshop	11, ABI	8, ABI	Intervention group: no change post intervention d= 0.11 ns	Intervention group: increased post intervention d= 1.0*	n/a	n/a	(-)
Tornas et al., 2016	Very good	RCT	GMT/ Brain Health workshop	31, ABI	35, ABI	Intervention group: increased post d= 0.48*	Intervention group: decreased post intervention d= 0.24 ns	n/a	n/a	(+)
Pooled Data				Total n= 74			No of tasks: 1/2 (50%) No of tasks: 1/2 (50%)		Inconsistent ± Inconsistent ±	

Abbreviations: No= number; Time Dev= Time deviation; GD= Garage Doors RCT= Randomised Control Trial; AIM= Assisted Intervention Monitoring; ABI= Acquired Brain Injury; n/a= not applicable; GMT= Goal Management Training; G= group; ns= non-significant.

Note: [a](+) sufficient, (?) indeterminate, (-) insufficient, or (?) mixed; [b]Unclear which Hotel score is reported in Gracey et al., 2017 and no available information to calculate within-group effect size ; * $p \leq .05$.

Appendix 2.1 Excel Hotel time deviation norms calculator

https://osf.io/v83hc/

Appendix 2.2 Normative data

https://osf.io/2wkzv/

Appendix 2.3 Score transformation Calculator

https://osf.io/9qe46/

Appendix 2.4 Ethical Approval Letter

Professor Jonathan Evans

MVLS College Ethics Committee

The Hotel task: Normative data on a measure of multitasking for the UK adult population

200200148

The College Ethics Committee has reviewed your application and has agreed that there is no objection on ethical grounds to the proposed study.
As you are using a data resource that already has consent and approvals for unspecified research by collaborators, arguably additional local approval is not required. However, we recognise that local review is often recommended for student projects.
We would ask that you follow the standard rules for a research study as outlined below.
In addition to ethical considerations, where there is transfer of data, there can be the need for additional approvals or a material transfer agreement. The data office can advise on this aspect of your project.

- The data should be held securely for a period of ten years after the completion of the research project, or for longer if specified by the research funder or sponsor, in accordance with the University's Code of Good Practice in Research: (http://www.gla.ac.uk/media/media_227599_en.pdf)

- The research should be carried out only on the sites, and/or groups defined in the application.

- Project end date as stipulated in original application.

- Any proposed changes in the protocol should be submitted for reassessment, except when it is necessary to change the protocol to eliminate hazard to the subjects or where the change involves only the administrative aspects of the project. The Ethics Committee should be informed of any such changes.

- You should submit a short end of study report within 3 months of completion.

Yours sincerely

Dr Terry Quinn

Terry Quinn
FWSO, FESO, MD, FRCP, BSc (hons), MBChB (hons)
Senior Lecturer / Honorary Consultant

College of Medicine, Veterinary & Life Sciences
Institute of Cardiovascular and Medical Sciences
New Lister Building, Glasgow Royal Infirmary
Glasgow G31 2ER
terry.quinn@glasgow.gla.ac.uk
Tel – 0141 201 8519

Appendix 2.5 Regression Output

RegBuild_MR_Raw.exe: Builds a multiple regression equation and uses it to make inferences concerning a case

THIS PROGRAM IMPLEMENTS STATISTICAL METHODS DEVELOPED IN THE FOLLOWING PAPER:
Crawford, J.R., Garthwaite, P.H., Denham, A.K., & Chelune, G.J. (2012). Using regression equations
built from summary data in the psychological assessment of the individual case: Extension to multiple
regression. Psychological Assessment, 24, 801-814. (doi: 10.1037/a0027699).

INPUTS :
Number of predictor (i.e., X) variables = 4
Sample size (n) for sample providing the summary data = 499
Credible limit required: Two-sided

Raw data for controls:

[1]:	14.20000	-22.00000	484.00000	20.00000	3.00000
[2]:	14.37000	-30.00000	900.00000	13.00000	13.00000
[3]:	14.45000	-36.00000	1296.00000	12.00000	15.00000
[4]:	11.91000	-30.00000	900.00000	13.00000	12.00000
[5]:	17.30000	-32.00000	1024.00000	17.00000	11.00000
[6]:	16.09000	-26.00000	676.00000	17.00000	16.00000
[7]:	16.46000	-26.00000	676.00000	20.00000	15.00000
[8]:	22.39000	-31.00000	961.00000	12.00000	6.00000
[9]:	12.62000	-31.00000	961.00000	16.00000	14.00000
[10]:	19.75000	-32.00000	1024.00000	16.00000	15.00000
[11]:	12.62000	-28.00000	784.00000	20.00000	15.00000
[12]:	11.51000	-36.00000	1296.00000	13.00000	11.00000
[13]:	17.24000	-28.00000	784.00000	13.00000	14.00000
[14]:	5.27000	-26.00000	676.00000	17.00000	14.00000
[15]:	10.71000	-28.00000	784.00000	19.00000	15.00000
[16]:	14.59000	-29.00000	841.00000	16.00000	12.00000
[17]:	9.71000	-27.00000	729.00000	17.00000	13.00000
[18]:	16.11000	-27.00000	729.00000	17.00000	17.00000
[19]:	14.10000	-28.00000	784.00000	20.00000	15.00000
[20]:	17.01000	-28.00000	784.00000	17.00000	16.00000
[21]:	17.43000	-34.00000	1156.00000	14.00000	9.00000
[22]:	12.00000	-29.00000	841.00000	18.00000	13.00000
[23]:	13.74000	-27.00000	729.00000	18.00000	10.00000
[24]:	20.82000	-27.00000	729.00000	16.00000	14.00000
[25]:	19.69000	-30.00000	900.00000	13.00000	13.00000
[26]:	9.95000	-23.00000	529.00000	12.00000	15.00000
[27]:	21.91000	-19.00000	361.00000	18.00000	8.00000
[28]:	16.36000	-18.00000	324.00000	17.00000	12.00000
[29]:	20.23000	-19.00000	361.00000	18.00000	13.00000

[30]:	22.52000	-24.00000	576.00000	11.00000	9.00000
[31]:	11.57000	-22.00000	484.00000	12.00000	10.00000
[32]:	20.30000	-24.00000	576.00000	19.00000	16.00000
[33]:	11.06000	-23.00000	529.00000	17.00000	12.00000
[34]:	14.96000	-25.00000	625.00000	17.00000	11.00000
[35]:	26.83000	-20.00000	400.00000	11.00000	11.00000
[36]:	7.66000	-20.00000	400.00000	17.00000	14.00000
[37]:	14.62000	-25.00000	625.00000	16.00000	13.00000
[38]:	26.83000	-17.00000	289.00000	20.00000	5.00000
[39]:	21.91000	-20.00000	400.00000	17.00000	14.00000
[40]:	17.92000	-17.00000	289.00000	16.00000	13.00000
[41]:	21.91000	-25.00000	625.00000	21.00000	14.00000
[42]:	14.07000	-20.00000	400.00000	16.00000	9.00000
[43]:	12.32000	-25.00000	625.00000	17.00000	16.00000
[44]:	16.28000	-23.00000	529.00000	13.00000	13.00000
[45]:	15.03000	-24.00000	576.00000	21.00000	18.00000
[46]:	8.95000	-17.00000	289.00000	20.00000	16.00000
[47]:	12.04000	-20.00000	400.00000	13.00000	5.00000
[48]:	15.73000	-21.00000	441.00000	19.00000	4.00000
[49]:	18.47000	-18.00000	324.00000	19.00000	11.00000
[50]:	17.50000	-19.00000	361.00000	14.00000	11.00000
[51]:	21.91000	-19.00000	361.00000	17.00000	10.00000
[52]:	15.25000	-21.00000	441.00000	17.00000	12.00000
[53]:	19.82000	-20.00000	400.00000	16.00000	16.00000
[54]:	9.64000	-25.00000	625.00000	18.00000	9.00000
[55]:	7.98000	-17.00000	289.00000	13.00000	12.00000
[56]:	18.39000	-23.00000	529.00000	20.00000	15.00000
[57]:	13.84000	-19.00000	361.00000	16.00000	5.00000
[58]:	24.29000	-22.00000	484.00000	13.00000	8.00000
[59]:	13.35000	-24.00000	576.00000	20.00000	14.00000
[60]:	15.69000	-25.00000	625.00000	20.00000	15.00000
[61]:	8.12000	-26.00000	676.00000	19.00000	3.00000
[62]:	21.69000	-22.00000	484.00000	19.00000	12.00000
[63]:	17.46000	-17.00000	289.00000	20.00000	13.00000
[64]:	12.93000	-22.00000	484.00000	20.00000	14.00000
[65]:	15.92000	-18.00000	324.00000	21.00000	15.00000
[66]:	11.92000	-19.00000	361.00000	16.00000	16.00000
[67]:	12.02000	-21.00000	441.00000	17.00000	9.00000
[68]:	16.94000	-21.00000	441.00000	17.00000	8.00000
[69]:	15.49000	-17.00000	289.00000	19.00000	15.00000
[70]:	7.93000	-19.00000	361.00000	16.00000	12.00000
[71]:	17.05000	-20.00000	400.00000	18.00000	13.00000
[72]:	11.16000	-17.00000	289.00000	13.00000	15.00000
[73]:	15.19000	-16.00000	256.00000	18.00000	11.00000
[74]:	10.08000	-16.00000	256.00000	17.00000	11.00000
[75]:	15.08000	-26.00000	676.00000	16.00000	8.00000
[76]:	19.59000	-25.00000	625.00000	17.00000	14.00000
[77]:	15.54000	-18.00000	324.00000	13.00000	14.00000
[78]:	17.78000	-23.00000	529.00000	17.00000	15.00000
[79]:	16.19000	-19.00000	361.00000	16.00000	14.00000

[80]:	11.91000	-25.00000	625.00000	17.00000	10.00000
[81]:	11.74000	-17.00000	289.00000	18.00000	12.00000
[82]:	8.38000	-18.00000	324.00000	17.00000	12.00000
[83]:	10.43000	-20.00000	400.00000	17.00000	13.00000
[84]:	15.72000	-17.00000	289.00000	16.00000	12.00000
[85]:	17.63000	-25.00000	625.00000	16.00000	16.00000
[86]:	6.86000	-19.00000	361.00000	17.00000	11.00000
[87]:	18.89000	-18.00000	324.00000	16.00000	14.00000
[88]:	18.51000	-20.00000	400.00000	13.00000	9.00000
[89]:	21.05000	-21.00000	441.00000	20.00000	13.00000
[90]:	9.18000	-16.00000	256.00000	17.00000	12.00000
[91]:	19.75000	-21.00000	441.00000	15.00000	13.00000
[92]:	14.36000	-24.00000	576.00000	12.00000	14.00000
[93]:	11.80000	-18.00000	324.00000	16.00000	14.00000
[94]:	23.59000	-20.00000	400.00000	11.00000	11.00000
[95]:	15.77000	-15.00000	225.00000	17.00000	11.00000
[96]:	14.77000	-9.00000	81.00000	11.00000	12.00000
[97]:	10.31000	-9.00000	81.00000	16.00000	14.00000
[98]:	10.55000	-15.00000	225.00000	16.00000	14.00000
[99]:	20.21000	-8.00000	64.00000	11.00000	13.00000
[100]:	15.68000	-9.00000	81.00000	16.00000	14.00000
[101]:	6.75000	-12.00000	144.00000	22.00000	16.00000
[102]:	18.83000	-11.00000	121.00000	14.00000	8.00000
[103]:	23.54000	-13.00000	169.00000	19.00000	8.00000
[104]:	21.91000	-14.00000	196.00000	16.00000	14.00000
[105]:	8.78000	-13.00000	169.00000	19.00000	12.00000
[106]:	11.02000	-9.00000	81.00000	14.00000	15.00000
[107]:	22.31000	-10.00000	100.00000	11.00000	8.00000
[108]:	6.22000	-7.00000	49.00000	20.00000	16.00000
[109]:	16.59000	-11.00000	121.00000	17.00000	12.00000
[110]:	11.81000	-7.00000	49.00000	13.00000	11.00000
[111]:	15.13000	-7.00000	49.00000	18.00000	14.00000
[112]:	24.67000	-7.00000	49.00000	17.00000	16.00000
[113]:	16.78000	-7.00000	49.00000	16.00000	12.00000
[114]:	11.80000	-7.00000	49.00000	16.00000	11.00000
[115]:	9.76000	-12.00000	144.00000	14.00000	12.00000
[116]:	11.04000	-14.00000	196.00000	15.00000	14.00000
[117]:	12.30000	-14.00000	196.00000	20.00000	13.00000
[118]:	12.18000	-14.00000	196.00000	19.00000	13.00000
[119]:	21.91000	-13.00000	169.00000	13.00000	11.00000
[120]:	15.83000	-11.00000	121.00000	12.00000	12.00000
[121]:	22.23000	-14.00000	196.00000	22.00000	14.00000
[122]:	20.01000	-10.00000	100.00000	12.00000	11.00000
[123]:	13.25000	-13.00000	169.00000	19.00000	12.00000
[124]:	9.72000	-6.00000	36.00000	16.00000	16.00000
[125]:	12.09000	-15.00000	225.00000	17.00000	12.00000
[126]:	17.77000	-14.00000	196.00000	20.00000	16.00000
[127]:	22.17000	-8.00000	64.00000	15.00000	16.00000
[128]:	10.16000	-10.00000	100.00000	16.00000	12.00000
[129]:	10.30000	-8.00000	64.00000	18.00000	15.00000

[130]:	12.47000	-9.00000	81.00000	13.00000	8.00000
[131]:	20.59000	-12.00000	144.00000	20.00000	12.00000
[132]:	24.52000	-6.00000	36.00000	16.00000	12.00000
[133]:	9.74000	-9.00000	81.00000	16.00000	14.00000
[134]:	21.91000	-12.00000	144.00000	17.00000	16.00000
[135]:	13.51000	-15.00000	225.00000	14.00000	12.00000
[136]:	14.96000	-15.00000	225.00000	17.00000	9.00000
[137]:	9.82000	-10.00000	100.00000	17.00000	12.00000
[138]:	4.49000	-10.00000	100.00000	11.00000	10.00000
[139]:	11.06000	-10.00000	100.00000	18.00000	9.00000
[140]:	20.03000	-11.00000	121.00000	17.00000	14.00000
[141]:	14.54000	-9.00000	81.00000	17.00000	14.00000
[142]:	23.21000	-8.00000	64.00000	11.00000	12.00000
[143]:	13.41000	-5.00000	25.00000	16.00000	14.00000
[144]:	10.05000	-12.00000	144.00000	19.00000	14.00000
[145]:	9.24000	-7.00000	49.00000	14.00000	11.00000
[146]:	16.18000	-3.00000	9.00000	13.00000	12.00000
[147]:	12.46000	1.00000	1.00000	13.00000	16.00000
[148]:	10.19000	1.00000	1.00000	17.00000	16.00000
[149]:	18.68000	3.00000	9.00000	12.00000	10.00000
[150]:	4.98000	3.00000	9.00000	20.00000	15.00000
[151]:	15.49000	3.00000	9.00000	19.00000	16.00000
[152]:	21.91000	0.00000	0.00000	13.00000	16.00000
[153]:	20.75000	-5.00000	25.00000	11.00000	5.00000
[154]:	16.98000	3.00000	9.00000	17.00000	16.00000
[155]:	12.85000	1.00000	1.00000	14.00000	12.00000
[156]:	26.83000	3.00000	9.00000	16.00000	11.00000
[157]:	20.27000	-2.00000	4.00000	16.00000	15.00000
[158]:	26.83000	4.00000	16.00000	21.00000	13.00000
[159]:	13.94000	-5.00000	25.00000	19.00000	15.00000
[160]:	21.91000	1.00000	1.00000	10.00000	12.00000
[161]:	14.76000	3.00000	9.00000	17.00000	16.00000
[162]:	15.49000	1.00000	1.00000	18.00000	16.00000
[163]:	19.64000	-1.00000	1.00000	18.00000	9.00000
[164]:	18.52000	-2.00000	4.00000	16.00000	12.00000
[165]:	15.45000	1.00000	1.00000	16.00000	16.00000
[166]:	21.91000	2.00000	4.00000	11.00000	12.00000
[167]:	14.27000	2.00000	4.00000	17.00000	13.00000
[168]:	21.91000	0.00000	0.00000	19.00000	12.00000
[169]:	24.15000	-1.00000	1.00000	13.00000	13.00000
[170]:	17.91000	-4.00000	16.00000	16.00000	12.00000
[171]:	11.51000	-5.00000	25.00000	14.00000	10.00000
[172]:	19.20000	4.00000	16.00000	13.00000	14.00000
[173]:	16.78000	4.00000	16.00000	13.00000	16.00000
[174]:	19.80000	4.00000	16.00000	15.00000	16.00000
[175]:	10.29000	-3.00000	9.00000	14.00000	16.00000
[176]:	16.50000	0.00000	0.00000	17.00000	1.00000
[177]:	21.45000	-6.00000	36.00000	11.00000	11.00000
[178]:	19.49000	3.00000	9.00000	16.00000	14.00000
[179]:	13.68000	2.00000	4.00000	16.00000	11.00000

[180]:	22.63000	-2.00000	4.00000	16.00000	11.00000
[181]:	18.91000	0.00000	0.00000	14.00000	2.00000
[182]:	21.91000	0.00000	0.00000	13.00000	16.00000
[183]:	10.24000	-5.00000	25.00000	18.00000	12.00000
[184]:	26.83000	-3.00000	9.00000	11.00000	12.00000
[185]:	26.57000	3.00000	9.00000	16.00000	11.00000
[186]:	13.09000	-2.00000	4.00000	17.00000	12.00000
[187]:	13.80000	-2.00000	4.00000	12.00000	11.00000
[188]:	22.85000	-6.00000	36.00000	16.00000	13.00000
[189]:	23.29000	-2.00000	4.00000	13.00000	8.00000
[190]:	14.26000	2.00000	4.00000	13.00000	12.00000
[191]:	21.17000	2.00000	4.00000	11.00000	12.00000
[192]:	19.24000	2.00000	4.00000	10.00000	13.00000
[193]:	14.03000	-1.00000	1.00000	12.00000	13.00000
[194]:	7.14000	2.00000	4.00000	19.00000	14.00000
[195]:	15.03000	-2.00000	4.00000	16.00000	13.00000
[196]:	19.05000	4.00000	16.00000	16.00000	16.00000
[197]:	21.91000	-2.00000	4.00000	13.00000	15.00000
[198]:	12.02000	-3.00000	9.00000	18.00000	13.00000
[199]:	13.17000	-5.00000	25.00000	17.00000	14.00000
[200]:	17.10000	2.00000	4.00000	12.00000	10.00000
[201]:	26.36000	-2.00000	4.00000	19.00000	13.00000
[202]:	12.19000	1.00000	1.00000	13.00000	15.00000
[203]:	9.26000	4.00000	16.00000	11.00000	11.00000
[204]:	7.96000	-3.00000	9.00000	11.00000	10.00000
[205]:	19.40000	-1.00000	1.00000	12.00000	12.00000
[206]:	22.97000	-4.00000	16.00000	21.00000	12.00000
[207]:	15.33000	1.00000	1.00000	11.00000	11.00000
[208]:	16.01000	-2.00000	4.00000	11.00000	11.00000
[209]:	20.59000	0.00000	0.00000	13.00000	13.00000
[210]:	17.85000	-1.00000	1.00000	16.00000	12.00000
[211]:	14.85000	1.00000	1.00000	16.00000	16.00000
[212]:	8.53000	-5.00000	25.00000	17.00000	14.00000
[213]:	14.96000	-3.00000	9.00000	13.00000	10.00000
[214]:	10.53000	-4.00000	16.00000	13.00000	14.00000
[215]:	17.48000	0.00000	0.00000	16.00000	12.00000
[216]:	11.21000	2.00000	4.00000	20.00000	16.00000
[217]:	13.01000	-5.00000	25.00000	16.00000	13.00000
[218]:	21.91000	-2.00000	4.00000	13.00000	12.00000
[219]:	14.05000	3.00000	9.00000	18.00000	14.00000
[220]:	14.86000	0.00000	0.00000	11.00000	9.00000
[221]:	17.65000	-6.00000	36.00000	13.00000	10.00000
[222]:	12.54000	-3.00000	9.00000	16.00000	16.00000
[223]:	18.33000	-1.00000	1.00000	11.00000	8.00000
[224]:	12.32000	-5.00000	25.00000	18.00000	12.00000
[225]:	13.60000	-2.00000	4.00000	11.00000	11.00000
[226]:	21.91000	-3.00000	9.00000	17.00000	12.00000
[227]:	18.10000	-1.00000	1.00000	13.00000	14.00000
[228]:	17.66000	0.00000	0.00000	15.00000	11.00000
[229]:	16.75000	1.00000	1.00000	16.00000	11.00000

[230]:	14.35000	1.00000	1.00000	16.00000	13.00000
[231]:	14.37000	-3.00000	9.00000	16.00000	15.00000
[232]:	18.11000	-4.00000	16.00000	16.00000	14.00000
[233]:	13.09000	0.00000	0.00000	13.00000	13.00000
[234]:	12.74000	-5.00000	25.00000	18.00000	14.00000
[235]:	11.13000	-2.00000	4.00000	17.00000	14.00000
[236]:	20.53000	-3.00000	9.00000	11.00000	14.00000
[237]:	14.66000	-4.00000	16.00000	20.00000	15.00000
[238]:	20.12000	2.00000	4.00000	13.00000	15.00000
[239]:	18.91000	-5.00000	25.00000	22.00000	13.00000
[240]:	16.01000	-4.00000	16.00000	13.00000	9.00000
[241]:	16.89000	6.00000	36.00000	11.00000	12.00000
[242]:	15.49000	4.00000	16.00000	9.50000	11.00000
[243]:	10.47000	5.00000	25.00000	17.00000	16.00000
[244]:	8.36000	8.00000	64.00000	13.00000	10.00000
[245]:	21.95000	11.00000	121.00000	11.00000	9.00000
[246]:	21.91000	8.00000	64.00000	13.00000	11.00000
[247]:	15.49000	4.00000	16.00000	12.00000	16.00000
[248]:	13.07000	10.00000	100.00000	10.00000	11.00000
[249]:	16.42000	12.00000	144.00000	16.00000	16.00000
[250]:	11.63000	12.00000	144.00000	20.00000	14.00000
[251]:	17.14000	13.00000	169.00000	17.00000	12.00000
[252]:	18.88000	12.00000	144.00000	18.00000	15.00000
[253]:	14.72000	6.00000	36.00000	17.00000	13.00000
[254]:	24.00000	6.00000	36.00000	11.00000	6.00000
[255]:	15.49000	8.00000	64.00000	16.00000	10.00000
[256]:	12.88000	10.00000	100.00000	16.00000	15.00000
[257]:	15.20000	5.00000	25.00000	16.00000	14.00000
[258]:	20.69000	6.00000	36.00000	14.00000	11.00000
[259]:	10.71000	7.00000	49.00000	11.00000	7.00000
[260]:	14.80000	12.00000	144.00000	11.00000	10.00000
[261]:	17.20000	10.00000	100.00000	12.00000	13.00000
[262]:	15.59000	10.00000	100.00000	16.00000	14.00000
[263]:	21.09000	10.00000	100.00000	16.00000	16.00000
[264]:	11.10000	9.00000	81.00000	10.00000	7.00000
[265]:	16.84000	8.00000	64.00000	11.00000	8.00000
[266]:	22.03000	11.00000	121.00000	11.00000	6.00000
[267]:	18.97000	13.00000	169.00000	12.00000	13.00000
[268]:	21.78000	14.00000	196.00000	18.00000	7.00000
[269]:	26.83000	14.00000	196.00000	10.00000	9.00000
[270]:	16.71000	10.00000	100.00000	16.00000	13.00000
[271]:	19.58000	7.00000	49.00000	13.00000	11.00000
[272]:	22.18000	11.00000	121.00000	17.00000	12.00000
[273]:	26.83000	8.00000	64.00000	17.00000	13.00000
[274]:	8.46000	11.00000	121.00000	13.00000	16.00000
[275]:	26.27000	12.00000	144.00000	11.00000	16.00000
[276]:	16.55000	14.00000	196.00000	19.00000	14.00000
[277]:	14.57000	6.00000	36.00000	16.00000	14.00000
[278]:	11.55000	8.00000	64.00000	13.00000	11.00000
[279]:	17.26000	7.00000	49.00000	11.00000	14.00000

[280]:	12.15000	6.00000	36.00000	20.00000	14.00000
[281]:	13.14000	5.00000	25.00000	11.00000	4.00000
[282]:	20.76000	5.00000	25.00000	11.00000	11.00000
[283]:	15.49000	8.00000	64.00000	11.00000	12.00000
[284]:	17.07000	7.00000	49.00000	17.00000	14.00000
[285]:	18.31000	9.00000	81.00000	13.00000	16.00000
[286]:	17.62000	10.00000	100.00000	17.00000	16.00000
[287]:	14.49000	6.00000	36.00000	17.00000	16.00000
[288]:	13.23000	13.00000	169.00000	12.00000	12.00000
[289]:	12.67000	4.00000	16.00000	19.00000	15.00000
[290]:	12.03000	10.00000	100.00000	14.00000	14.00000
[291]:	15.88000	7.00000	49.00000	16.00000	15.00000
[292]:	15.76000	6.00000	36.00000	17.00000	16.00000
[293]:	17.35000	11.00000	121.00000	17.00000	12.00000
[294]:	11.90000	8.00000	64.00000	10.00000	13.00000
[295]:	17.75000	10.00000	100.00000	16.00000	14.00000
[296]:	14.09000	13.00000	169.00000	22.00000	11.00000
[297]:	18.12000	12.00000	144.00000	18.00000	16.00000
[298]:	6.20000	13.00000	169.00000	17.00000	15.00000
[299]:	16.29000	11.00000	121.00000	11.00000	14.00000
[300]:	12.68000	7.00000	49.00000	16.00000	9.00000
[301]:	12.85000	5.00000	25.00000	16.00000	13.00000
[302]:	22.96000	7.00000	49.00000	16.00000	12.00000
[303]:	21.91000	5.00000	25.00000	19.00000	14.00000
[304]:	22.04000	13.00000	169.00000	10.00000	9.00000
[305]:	17.46000	12.00000	144.00000	22.00000	16.00000
[306]:	14.16000	9.00000	81.00000	17.00000	14.00000
[307]:	26.54000	9.00000	81.00000	13.00000	12.00000
[308]:	14.18000	11.00000	121.00000	12.00000	16.00000
[309]:	19.46000	9.00000	81.00000	16.00000	16.00000
[310]:	19.21000	4.00000	16.00000	16.00000	13.00000
[311]:	18.65000	12.00000	144.00000	17.00000	14.00000
[312]:	13.14000	14.00000	196.00000	10.00000	9.00000
[313]:	14.98000	5.00000	25.00000	19.00000	16.00000
[314]:	15.59000	9.00000	81.00000	17.00000	13.00000
[315]:	22.23000	11.00000	121.00000	16.00000	13.00000
[316]:	18.70000	10.00000	100.00000	18.00000	16.00000
[317]:	13.13000	7.00000	49.00000	19.00000	13.00000
[318]:	15.49000	5.00000	25.00000	11.00000	12.00000
[319]:	26.59000	9.00000	81.00000	13.00000	15.00000
[320]:	11.05000	10.00000	100.00000	19.00000	16.00000
[321]:	11.46000	12.00000	144.00000	16.00000	15.00000
[322]:	8.49000	6.00000	36.00000	13.00000	14.00000
[323]:	12.84000	12.00000	144.00000	22.00000	14.00000
[324]:	15.50000	11.00000	121.00000	16.00000	11.00000
[325]:	16.40000	10.00000	100.00000	16.00000	16.00000
[326]:	21.91000	12.00000	144.00000	11.00000	11.00000
[327]:	9.57000	12.00000	144.00000	13.00000	15.00000
[328]:	15.14000	5.00000	25.00000	11.00000	10.00000
[329]:	20.83000	7.00000	49.00000	10.00000	9.00000

[330]:	26.83000	13.00000	169.00000	11.00000	12.00000
[331]:	19.16000	14.00000	196.00000	9.00000	4.00000
[332]:	11.74000	9.00000	81.00000	16.00000	15.00000
[333]:	14.87000	14.00000	196.00000	18.00000	16.00000
[334]:	21.91000	14.00000	196.00000	16.00000	14.00000
[335]:	21.91000	16.00000	256.00000	11.50000	9.00000
[336]:	21.96000	18.00000	324.00000	12.00000	11.00000
[337]:	19.52000	18.00000	324.00000	10.00000	12.00000
[338]:	16.11000	17.00000	289.00000	17.00000	12.00000
[339]:	15.03000	14.00000	196.00000	18.00000	14.00000
[340]:	21.59000	24.00000	576.00000	16.00000	10.00000
[341]:	20.20000	19.00000	361.00000	10.00000	12.00000
[342]:	11.28000	22.00000	484.00000	16.00000	16.00000
[343]:	18.01000	15.00000	225.00000	15.00000	12.00000
[344]:	12.48000	23.00000	529.00000	12.50000	13.00000
[345]:	19.09000	15.00000	225.00000	16.00000	16.00000
[346]:	19.87000	14.00000	196.00000	12.00000	12.00000
[347]:	15.96000	22.00000	484.00000	22.00000	15.00000
[348]:	21.91000	18.00000	324.00000	11.00000	9.00000
[349]:	12.91000	21.00000	441.00000	16.00000	13.00000
[350]:	16.27000	23.00000	529.00000	12.00000	16.00000
[351]:	18.98000	22.00000	484.00000	11.00000	15.00000
[352]:	21.91000	15.00000	225.00000	21.00000	12.00000
[353]:	17.34000	18.00000	324.00000	13.00000	8.00000
[354]:	19.62000	15.00000	225.00000	13.00000	1.00000
[355]:	25.48000	17.00000	289.00000	12.00000	10.00000
[356]:	21.92000	16.00000	256.00000	20.00000	16.00000
[357]:	6.56000	21.00000	441.00000	11.00000	11.00000
[358]:	13.07000	18.00000	324.00000	12.00000	14.00000
[359]:	21.82000	22.00000	484.00000	14.00000	12.00000
[360]:	23.40000	19.00000	361.00000	11.00000	1.00000
[361]:	26.83000	16.00000	256.00000	10.00000	5.00000
[362]:	26.30000	23.00000	529.00000	10.00000	5.00000
[363]:	22.73000	22.00000	484.00000	11.00000	16.00000
[364]:	16.25000	17.00000	289.00000	10.00000	8.00000
[365]:	14.40000	18.00000	324.00000	11.00000	7.00000
[366]:	16.41000	15.00000	225.00000	11.00000	3.00000
[367]:	19.09000	20.00000	400.00000	16.00000	15.00000
[368]:	9.88000	16.00000	256.00000	10.00000	12.00000
[369]:	22.95000	23.00000	529.00000	13.00000	16.00000
[370]:	22.98000	17.00000	289.00000	10.00000	11.00000
[371]:	13.90000	24.00000	576.00000	18.00000	16.00000
[372]:	15.67000	16.00000	256.00000	17.00000	14.00000
[373]:	21.93000	20.00000	400.00000	13.00000	14.00000
[374]:	17.88000	16.00000	256.00000	16.00000	15.00000
[375]:	13.14000	18.00000	324.00000	22.00000	16.00000
[376]:	18.50000	15.00000	225.00000	20.00000	13.00000
[377]:	20.61000	16.00000	256.00000	12.00000	9.00000
[378]:	26.83000	17.00000	289.00000	13.00000	3.00000
[379]:	16.54000	17.00000	289.00000	10.00000	13.00000

[380]:	10.74000	19.00000	361.00000	13.00000	13.00000
[381]:	10.25000	17.00000	289.00000	18.00000	16.00000
[382]:	16.11000	14.00000	196.00000	16.00000	16.00000
[383]:	14.91000	21.00000	441.00000	18.00000	14.00000
[384]:	26.83000	19.00000	361.00000	10.00000	12.00000
[385]:	16.92000	18.00000	324.00000	16.00000	10.00000
[386]:	17.50000	22.00000	484.00000	16.00000	16.00000
[387]:	24.30000	14.00000	196.00000	10.00000	12.00000
[388]:	19.50000	16.00000	256.00000	15.00000	14.00000
[389]:	18.22000	22.00000	484.00000	16.00000	13.00000
[390]:	16.66000	23.00000	529.00000	20.00000	16.00000
[391]:	14.83000	21.00000	441.00000	10.00000	12.00000
[392]:	11.95000	20.00000	400.00000	10.00000	9.00000
[393]:	22.37000	24.00000	576.00000	10.00000	11.00000
[394]:	13.87000	15.00000	225.00000	13.00000	4.00000
[395]:	17.31000	22.00000	484.00000	11.00000	9.00000
[396]:	18.19000	14.00000	196.00000	22.00000	16.00000
[397]:	23.74000	22.00000	484.00000	18.00000	14.00000
[398]:	12.87000	15.00000	225.00000	10.00000	1.00000
[399]:	17.03000	24.00000	576.00000	11.00000	12.00000
[400]:	22.75000	23.00000	529.00000	10.00000	16.00000
[401]:	16.38000	21.00000	441.00000	10.00000	7.00000
[402]:	20.01000	21.00000	441.00000	10.00000	10.00000
[403]:	16.91000	21.00000	441.00000	20.00000	16.00000
[404]:	12.99000	15.00000	225.00000	12.00000	14.00000
[405]:	19.46000	20.00000	400.00000	13.00000	15.00000
[406]:	24.77000	16.00000	256.00000	18.00000	11.00000
[407]:	8.76000	17.00000	289.00000	12.00000	16.00000
[408]:	24.53000	21.00000	441.00000	12.00000	15.00000
[409]:	17.70000	24.00000	576.00000	11.00000	14.00000
[410]:	24.65000	19.00000	361.00000	17.00000	16.00000
[411]:	21.91000	15.00000	225.00000	13.00000	14.00000
[412]:	21.48000	19.00000	361.00000	17.00000	16.00000
[413]:	17.23000	20.00000	400.00000	16.00000	16.00000
[414]:	25.67000	20.00000	400.00000	18.00000	14.00000
[415]:	19.51000	24.00000	576.00000	18.00000	16.00000
[416]:	20.89000	20.00000	400.00000	11.00000	14.00000
[417]:	15.53000	22.00000	484.00000	21.00000	14.00000
[418]:	12.04000	18.00000	324.00000	20.00000	16.00000
[419]:	14.33000	17.00000	289.00000	12.00000	14.00000
[420]:	17.33000	23.00000	529.00000	13.00000	16.00000
[421]:	16.07000	24.00000	576.00000	15.00000	12.00000
[422]:	11.54000	15.00000	225.00000	17.00000	16.00000
[423]:	11.20000	17.00000	289.00000	12.00000	15.00000
[424]:	20.07000	16.00000	256.00000	16.00000	15.00000
[425]:	22.33000	24.00000	576.00000	16.00000	15.00000
[426]:	21.19000	24.00000	576.00000	11.00000	9.00000
[427]:	11.48000	17.00000	289.00000	19.00000	16.00000
[428]:	19.80000	22.00000	484.00000	10.00000	11.00000
[429]:	21.91000	16.00000	256.00000	21.00000	15.00000

[430]:	21.91000	22.00000	484.00000	16.00000	16.00000
[431]:	17.70000	21.00000	441.00000	13.00000	16.00000
[432]:	14.74000	15.00000	225.00000	11.00000	11.00000
[433]:	22.80000	18.00000	324.00000	21.00000	13.00000
[434]:	20.89000	18.00000	324.00000	12.00000	14.00000
[435]:	10.63000	25.00000	625.00000	10.00000	13.00000
[436]:	21.91000	19.00000	361.00000	10.00000	10.00000
[437]:	26.83000	24.00000	576.00000	9.00000	9.00000
[438]:	22.23000	29.00000	841.00000	11.00000	13.00000
[439]:	17.43000	31.00000	961.00000	16.00000	13.00000
[440]:	26.62000	29.00000	841.00000	16.00000	15.00000
[441]:	16.20000	24.00000	576.00000	18.00000	11.00000
[442]:	16.73000	25.00000	625.00000	12.00000	16.00000
[443]:	23.25000	25.00000	625.00000	13.00000	16.00000
[444]:	16.13000	27.00000	729.00000	17.00000	15.00000
[445]:	21.91000	29.00000	841.00000	17.00000	14.00000
[446]:	21.91000	27.00000	729.00000	16.00000	12.00000
[447]:	25.99000	30.00000	900.00000	11.00000	16.00000
[448]:	17.18000	26.00000	676.00000	11.00000	16.00000
[449]:	22.74000	32.00000	1024.00000	9.00000	12.00000
[450]:	16.96000	28.00000	784.00000	13.00000	16.00000
[451]:	21.91000	25.00000	625.00000	9.00000	13.00000
[452]:	20.70000	26.00000	676.00000	9.00000	11.00000
[453]:	10.16000	29.00000	841.00000	10.00000	14.00000
[454]:	21.91000	31.00000	961.00000	13.00000	16.00000
[455]:	24.83000	29.00000	841.00000	11.00000	12.00000
[456]:	17.52000	30.00000	900.00000	16.00000	11.00000
[457]:	18.26000	28.00000	784.00000	9.00000	14.00000
[458]:	13.12000	30.00000	900.00000	14.00000	15.00000
[459]:	19.72000	25.00000	625.00000	9.00000	11.00000
[460]:	15.49000	25.00000	625.00000	11.00000	14.00000
[461]:	23.56000	26.00000	676.00000	9.00000	4.00000
[462]:	21.73000	34.00000	1156.00000	11.00000	16.00000
[463]:	17.63000	26.00000	676.00000	9.00000	5.00000
[464]:	18.23000	31.00000	961.00000	18.00000	16.00000
[465]:	24.81000	26.00000	676.00000	18.00000	15.00000
[466]:	24.18000	31.00000	961.00000	15.00000	12.00000
[467]:	21.91000	33.00000	1089.00000	10.00000	14.00000
[468]:	17.44000	28.00000	784.00000	9.00000	10.00000
[469]:	16.74000	28.00000	784.00000	17.00000	15.00000
[470]:	23.93000	26.00000	676.00000	9.00000	12.00000
[471]:	19.98000	25.00000	625.00000	19.00000	16.00000
[472]:	25.04000	25.00000	625.00000	16.00000	13.00000
[473]:	18.18000	24.00000	576.00000	20.00000	16.00000
[474]:	20.03000	26.00000	676.00000	12.00000	16.00000
[475]:	16.65000	30.00000	900.00000	16.00000	16.00000
[476]:	15.49000	26.00000	676.00000	11.00000	12.00000
[477]:	9.57000	26.00000	676.00000	19.00000	15.00000
[478]:	22.91000	26.00000	676.00000	13.00000	6.00000
[479]:	10.88000	29.00000	841.00000	13.00000	16.00000

[480]:	11.71000	32.00000	1024.00000	10.00000	12.00000
[481]:	17.76000	28.00000	784.00000	9.00000	12.00000
[482]:	19.18000	26.00000	676.00000	11.00000	10.00000
[483]:	15.15000	25.00000	625.00000	19.00000	15.00000
[484]:	10.25000	30.00000	900.00000	15.00000	16.00000
[485]:	17.99000	27.00000	729.00000	17.00000	16.00000
[486]:	8.50000	26.00000	676.00000	20.00000	16.00000
[487]:	14.23000	28.00000	784.00000	17.00000	12.00000
[488]:	12.46000	26.00000	676.00000	16.00000	16.00000
[489]:	8.94000	27.00000	729.00000	12.00000	13.00000
[490]:	26.36000	24.00000	576.00000	20.00000	14.00000
[491]:	26.83000	29.00000	841.00000	10.00000	11.00000
[492]:	21.91000	26.00000	676.00000	13.00000	16.00000
[493]:	12.52000	28.00000	784.00000	11.00000	13.00000
[494]:	26.83000	25.00000	625.00000	15.00000	14.00000
[495]:	11.75000	26.00000	676.00000	18.00000	16.00000
[496]:	14.63000	25.00000	625.00000	13.00000	11.00000
[497]:	12.36000	25.00000	625.00000	11.00000	15.00000
[498]:	23.73000	30.00000	900.00000	10.00000	16.00000
[499]:	21.91000	26.00000	676.00000	18.00000	16.00000

INPUTS: Summary statistics for sample providing the data together with scores of the case:

Measure	Control Mean	Control SD	Scores for Case
Criterion (Y) :	16.861	4.880	20.260
Predictor (X) 1 :	3.553	17.149	-6.000
Predictor (X) 2 :	306.126	289.012	36.000
Predictor (X) 3 :	14.937	3.313	17.000
Predictor (X) 4 :	12.627	3.051	14.000

INPUTS: Correlation(s) between criterion and predictor(s) in sample:

1.00000	0.24776	0.08274	-0.18969	-0.12235
0.24776	1.00000	0.11655	-0.30313	0.09359
0.08274	0.11655	1.00000	-0.07038	0.07298
-0.18969	-0.30313	-0.07038	1.00000	0.33678
-0.12235	0.09359	0.07298	0.33678	1.00000

OUTPUTS: Regression equation (alpha & betas) and squared semi-partial correlation for predictors:

Predictor	Beta	Squared semi-partial correlation
Intercept (alpha):	20.443	-
Predictor (X) 1 :	0.065	0.053
Predictor (X) 2 :	0.001	0.004
Predictor (X) 3 :	-0.110	0.006
Predictor (X) 4 :	-0.197	0.015

OUTPUTS: FURTHER RESULTS FOR THE MULTIPLE REGRESSION MODEL:
Standard error of estimate for regression equation = 4.672
Multiple R for regression equation = 0.302
R Squared for regression equation = 0.091
Adjusted (shrunken) R Squared for regression equation = 0.084
Significance test for overall regression: F [4, 494] = 12.3709
Significance test for overal regression: p value = 0.0000

OUTPUTS: RESULTS FROM ANALYSIS OF THE INDIVIDUAL CASE:

Case's OBTAINED score on Task of Interest = 20.2600
Case's PREDICTED score from regression equation = 15.4670
Discrepancy (obtained minus predicted) between case's obtained and predicted scores = 4.7930

Effect size (Z-OP) for discrepancy between obtained and predicted scores (plus 95% CI):
Effect size (Z-OP) = 1.030 (95% CI = 0.874 to 1.177)

Standard error for an additional (i.e., N + 1th) case = 4.6830

Significance test (t) on the discrepancy between the case's obtained and predicted scores:
t value (on 494 df) = 1.0235
One-tailed probability = 0.1533
Two-tailed probability = 0.3066

Estimated percentage of population obtaining a discrepancy more extreme than the case = 15.329250%
95% confidence limits on the percentage = 11.9565% to 19.1075%

Appendix 2.6 Original Final Approved MRP Proposal

https://osf.io/9dq6v/

Appendix 2.7 Hotel Test/MRP proposal

https://osf.io/6s97f/